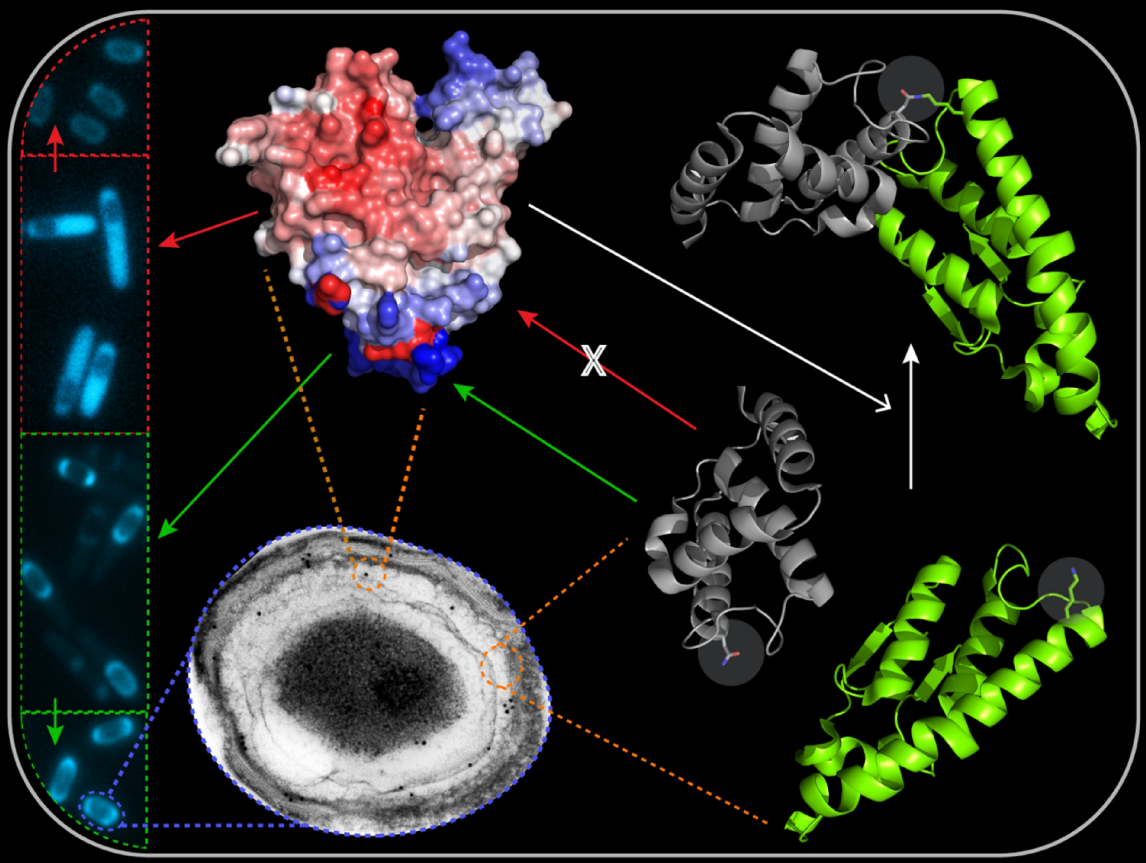
Assembly and function of a protein cross-linking enzyme during bacterial spore morphogenesis

Catarina G. Fernandes



Dissertation presented to obtain the Ph.D degree in Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
June, 2014

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Assembly and function of a protein cross-linking enzyme during bacterial spore morphogenesis

Catarina G. Fernandes

Dissertation presented to obtain the Ph.D degree in Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, June, 2014

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Nas partes desta Tese que foram escritas em Português não foi aplicado o novo acordo ortográfico.

ACKOWLEDGEMENTS

To Instituto de Tecnologia Química e Biológica (ITQB) of Universidade Nova de Lisboa for providing all the working conditions necessary for the development of this work.

To Fundação para a Ciência e Tecnologia (FCT) for the financial support through the individual PhD fellowship SFRH/BD/43200/2008, with which I was awarded.

To Adriano O. Henriques for accepting me in his lab. For the useful discussions and support throughout the years of my PhD. For his endless enthusiasm and constructive critics and for encouraging the open and free work environment within his group. Finally for his patience and readiness during the writing of this thesis.

To Anabela, who initially guided me in the laboratory work. Thank you for all you taught me, for the good sense of humor and for the incredible patience when I knew nothing.

To past and present members of the Microbial Development Group: Ana, João Bota, Mónica Serrano, Gonçalo, Teresa Costa, Patrícia, Carolina Freitas, Fátima, Wilson, Anabela and Filipa for making this group a fun place to work in; for the help with the little things and useful discussions and for the fun times outside the lab. To the lab neighbors: Raquel, Pedrinho, João, Matos, Magda, Helena, Bruno, Mafalda, Nathalie, Trish and James for their companionship, fun lunch and coffee times and for their willingness to help.

To Teresa Costa, Mónica Serrano, Cláudia and Matos for the little "tips" which simplified the lab work on numerous occasions and made my life so much easier.

To Mónica Serrano, Raquel and Inês for reading some of the chapters and for the important feedback. Inês, you were awesome; I cannot believe I've been writing traverse wrong for the last 6 years and only now I've been corrected.

A special thanks to Teresa Baptista da Silva, an incredibly competent and lovely person which made my life so much easier. I don't think you

understand how much your help was appreciated.

To dear friends: Inês, Matos, Claúdia, Raquel, Bruno, João, Pedrinho, Madga and Mónica. Thank you for the entertaining nights out, fun times and for making me laugh. Inês and Matos, longtime friends, thank you for the company and support in darker times. To Teresa Avelar, thank you for coming back into my life, for keeping me sane during the writing of the thesis and for making this last year so much more fun.

Finalmente, ao meus pais, sem os quais esta tese seria impossível. Obrigado por todo o apoio ao longo dos meus anos de existência e por me encoragem a seguir o meu caminho.

TABLE OF CONTENTS

ABSTRACT vii
RESUMO x

Chapter 1
General introduction 1
BACTERIAL ENDOSPORES 3
SPORULATION IN BACILLUS SUBTILIS 4
Structure and functional characteristics of the spore 4
Morphological stages coupled to gene expression regulation 7
Spore coat formation 8
Tgl 17
TRANSGLUTAMINASES 19
Animal TGases 19
Non-animal TGases 23
REFERENCES 25

Chapter 2
Structural and Functional Characterization of a Streamlined Bacterial Transglutaminase 35
ABSTRACT 37
INTRODUCTION 37
MATERIALS AND METHODS 41
RESULTS 48
Structure of Tgl 48
The hydrophobic tunnel of Tgl and substrate accessibility 50
The catalytic core of Tgl adopts the NlpC/P60 papain-like fold 52
Active site architecture 54
Probing the active site of Tgl 55
Conformation of the putative catalytic residues at pH 7 58
Interactions among residues near the catalytic Cys 61
A new active site residue 63
A partially redundant catalytic diad: non-reciprocal substitution of Glu187 by Glu115 . 63
DISCUSSION 65
REFERENCES 68

Chapter 3
Auto-regulation of SafA assembly via the subcellular localization of a protein cross-linking enzyme 75
ABSTRACT 77
INTRODUCTION 77

MATERIALS AND METHODS 81
RESULTS 91
An in-frame deletion mutant of safA 91
Dependency of SafA and C30 on Tgl 92
The extractability of cortex-associated SafA and C30 is dependent on Tgl 94
Tgl is also associated with the cortex layer of spores 98
Tgl localizes to the cortex, inner and outer coat layers 100
Construction of a functional tgl-cfp *fusion protein* 100
The assembly of Tgl during sporulation is dependent on SafA 105
CotE is necessary for Tgl to remain associated with free spores 109
The localization of Tgl-CFP is affected by the inhibition of cortex synthesis 109
DISCUSSION 111
REFERENCES 115

Chapter 4

The assembly of a sporulation transglutaminase is driven by its substrates 123
ABSTRACT 125
INTRODUCTION 125
MATERIALS AND METHODS 128
RESULTS 137
The substrates of Tgl appear to influence the recruitment and assembly of Tgl-CFP to the forespore 137
Residues at the entrance of the tunnel of Tgl are important for the overall activity of the enzyme 144
Front side residues of the tunnel of Tgl, likely involved in substrate interactions, are also important for the recruitment of the enzyme during sporulation 148
Residues important for the recruitment of Tgl to the spore are also important for the in vitro cross-linking of C30 152
C30 retains Tgl^{wt} in the mother cell but not Tgl mutants with altered front side residues 154
DISCUSSION 155
REFERENCES 162

Chapter 5

General discussion 167
TGL AS THE MINIMAL TGASE: IMPLICATIONS FOR TGL FUNCTION 170
The minimal features for the catalysis of protein cross-linking reactions 170
Activity regulation is coupled to substrate specificity 171
Impact on the localization mechanism of Tgl 172
Activity regulation by substrate alterations 172
HOW DOES TGL REACH THE CORTEX? 173
TGL CEMENTS THE CORTEX/INNER COAT INTERFACE 175
BACTERIAL TRANSGLUTAMINASES AND TGL 176

MTG 176
CNF1 and DNT 178
WbmE and TgpA 179
A global view of bacterial TGases (or TGase-like proteins) 180
REFERENCES 181

TABLE OF FIGURES

Chapter 1

General introduction

Figure 1.1. Structure of the spore and the sporulation process 6

Figure 1.2. Spore coat formation is controlled by a hierarchy of assembly dependencies. 9

Figure 1.3. Spore coat assembly is regulated by the spatial arrangement of the different coat layers and by successive waves of encasement 11

Figure 1.4. Interactions between morphogenetic proteins drive spore coat assembly 13

Figure 1.5. Cartoon representation of the inactive and active structures of Factor XIII/Factor XIIIa, TGase 2 and TGase 3 20

Chapter 2

Structural and Functional Characterization of a Streamlined Bacterial Transglutaminase

Figure 2.1. Proposed reaction mechanism for the animal TGases 39

Figure 2.2. Superimposition of the active site region of monomers A and B of the Tgl:cysteamine X-ray structure 49

Figure 2.3. Overall structure of Tgl 50

Figure 2.4. The Tgl tunnel 51

Figure 2.5. Conservation among the tunnel residues 52

Figure 2.6. The catalytic core of Tgl has an NlpC/P60-like fold 53

Figure 2.7. Assessment of the enzymatic activity of wild type and mutant forms of Tgl, by BSA labeling 56

Figure 2.8. Assessment of the enzymatic activity of wild type and mutant forms of Tgl, by GST-SpoVID202 cross-linking 57

Figure 2.9. Behaviour of His200 during MD simulations 60

Figure 2.10. Behavior of active site residues during MD simulations with Tgl^{E187A} and Tgl^{E115A} 64

Chapter 3

Auto-regulation of SafA assembly via the subcellular localization of a protein cross-linking enzyme

Figure 3.1. An in-frame *safA* deletion can be complemented to restore the wild type phenotype. 92

Figure 3.2. Confirmation of SafA and C30 as substrates of Tgl 93

Figure 3.3. Effect of the absence of Tgl on the extractability of SafA and C30 in the different layers of the spore 95

Figure 3.4. Tgl can be extracted from different layers of the spore 99

Figure 3.5. Tgl localizes mainly in the cortex and cortex/inner coat interface as detected by immunogold electron microscopy 100

Figure 3.6. *tgl-hl4-cfp* fails to complement a Δ*tgl* mutation 102

Figure 3.7. Production of Tgl-HL4-CFP from a consensus promoter fully complements a *tgl* deletion mutation 104

Figure 3.8. The proper assembly of Tgl is dependent on SafA, CotE and cortex formation 106

Figure 3.9. The proper assembly of Tgl is dependent on SafA, CotE and cortex formation (*cont.*) 107

Figure 3.10. The assembly of Tgl is organized in two steps, and the final assembly state of the enzyme and SafA in the cortex, appears to be interdependent 112

Chapter 4

The assembly of a sporulation transglutaminase is driven by its substrates

Figure 4.1. The assembly of Tgl appears to be regulated by its substrates 138

Figure 4.2. The assembly of Tgl appears to be regulated by its substrates (*cont.*) 140

Figure 4.3. Identification of tunnel entrance residues that may be important for substrate interactions 145

Figure 4.4. Residues at both entrances of the tunnel of Tgl are important for the enzymatic activity of the enzyme 147

Figure 4.5. The exchange of Trp149 to Ala alters the specificity of Tgl 148

Figure 4.6. Residues located at the front entrance of the tunnel of Tgl have a defining role in the recruitment of the enzyme to the forespore 150

Figure 4.7. Residues located at the front entrance of the tunnel of Tgl have a defining role in the recruitment of the enzyme to the forespore (*cont.*) 151

Figure 4.8. Non-catalytic residues, important for the recruitment of Tgl to the forespore during sporulation, are also important for the *in vitro* cross-linking of C30 ...153

Figure 4.9. Before cortex maturation, in Δ*safA*$_{FL}$, C30 traps Tgl-CFP in an erroneous localization ...154

Figure 4.10. The aggregation of *tgl-cfp* seen in Δ*safA*$_{FL}$ is lost in *tgl-cfp* mutants where front side residues were exchanged to Ala...156

Chapter 5

General discussion

Figure 5.1. The complementation of the *safA*$_{spc}$ mutant does not restore the cortex association of SafA and C30 ...175

Figure 5.2. Structural representation of the front side of Tgl, mature MTG, and C-terminal catalytic domain of CNF1 ...177

ABSTRACT

Sporulation in *Bacillus subtilis* culminates with the formation of a dormant endospore. The endospore (or spore) is one of the most resilient cell types known and can remain viable in the environment for extended periods of time. Contributing to the spore's resistance and its ability to interact with and monitor its immediate environment is the coat, the outermost layer of *B. subtilis* spores. The coat is composed by over 70 different proteins, which are produced at different stages in sporulation and orderly assembled around the developing spore. Central to the coat assembly pathway is a group of morphogenetic proteins, which guide the deposition of the coat proteins, and whose absence severely impairs the formation and functional properties of the coat. Nonetheless, several other processes appear to be important for proper coat formation including protein-protein cross-linking.

Tgl, a spore coat transglutaminase (TGase), catalyzes the cross-linking of at least four spore coat proteins. TGases are best known for their ability to catalyze the formation of ε-(γ-glutamyl)lysyl isopeptide bonds between the side chains of protein-bound glutaminyl and lysyl residues (Q and K substrates, respectively), resulting in protein cross-linking. TGases are often associated with the stabilization of tissues or macromolecular assemblies and play important roles in developmental processes. The structural analysis of TGases indicates that these enzymes along with papain-like cysteine proteases and peptidoglycan endopeptidases have evolved from a common ancestor. Although TGases appear to be widespread and present in the three domains of life, our current knowledge of this group of enzymes is almost exclusively derived from studies on the animal enzymes. To date, only one bacterial TGase, from a *Streptomyces* species, has been analyzed in detail. It deviates considerably from the animal TGase paradigm, suggesting that much is to be learned from further studies of bacterial TGases. Here, we have studied Tgl, with the overall goal of contributing to a better structural and mechanistic understanding of the bacterial TGases and in an attempt to relate its assembly and function at the spore surface to its molecular characteristics.

We first solved the structure of Tgl. We found that, while the overall fold

of the enzyme is unique, the catalytic core of Tgl is related to that of peptidoglycan cysteine endopeptidases, in turn thought to be the closest relative of the ancestral fold common to TGases and papain-like cysteine peptidases. These observations lend support to the view that at least some TGases have derived from an ancient ancestral, common to different families of enzymes. While animal TGases function through a conserved catalytic Cys-His-Asp triad, activity assays with wild type and mutant forms of Tgl indicate that the enzyme functions via a partially redundant Cys116-Glu187/Glu115 catalytic diad. The presence of a catalytic diad may prove to be a prevalent feature of bacterial TGases. Tgl also appears to be a stripped down TGase and may represent the minimal structural features necessary for the catalysis of protein cross-linking reactions. Our structural and biochemical characterization of Tgl shows that the enzyme is synthesized in active form, in striking contrast with all the other TGases characterized to date. Also striking is the presence of a hydrophobic tunnel that transverses the molecule from side to side with the active site located at its center. Such a tunnel is only exposed in the active forms of the animal TGases and it may have a central role in catalysis, by modulating the access of the substrates to the catalytic center.

We then analyzed the role of Tgl in spore formation. Surprisingly, immunogold labeling and fractionation experiments showed that Tgl, not only associates with the inner coat but also with the cortex layer of the endospore. The cortex is composed by a modified form of peptidoglycan and is located inwards and juxtaposed to the inner coat. SafA is a morphogenetic protein that controls assembly of the inner spore coat. SafA and its short form C30, two spore coat substrates of Tgl, are also located in the inner coat and cortex layers and their extractability from both layers is controlled by Tgl. Conversely, deletion of *safA* reduces the levels of Tgl in the coat and results in its absence from the cortex layer. Together, these results show that the assembly of both Tgl and SafA is controlled by an interdependent auto-regulatory loop, and suggest that Tgl has a role in cementing the interface between the cortex and coat layers of the spore. Additionally, using a functional Tgl-CFP fusion protein we determined that *safA* is important for the correct localization of Tgl-CFP during sporulation.

Nonetheless, a significant fraction of Tgl still displays an apparently correct localization pattern in the absence of SafA. This implies that the assembly of Tgl relies on redundant interactions.

Because SafA is a substrate of Tgl we investigated whether other substrates contributed to the assembly of the enzyme. We found that the absence of the known substrates of Tgl affects, to different degrees, the recruitment of Tgl-CFP to the spore during sporulation. Furthermore, delocalization of C30, one of the substrates of Tgl, also delocalizes Tgl-CFP in a similar pattern. Finally, we identified non-catalytic residues in Tgl, presumed to be involved in substrate interactions. The individual substitution of these residues to Ala impairs the recruitment of Tgl-CFP to the spore surface during sporulation. Thus our results strongly suggest that the assembly of Tgl is dependent on the assembly of its substrates. The Tgl tunnel segregates the docking sites of the Q and K substrates to its opposite entrances. We have identified the Q and K sides of Tgl and we show that residues located at both entrances of the tunnel as well as enzymatic activity are important for the recruitment of Tgl to the forespore. However, only Q side residues appear to be paramount for the assembly of the enzyme. This observation, in particular, suggests that the formation of covalent Q-substrate-enzyme intermediates (corresponding to the first step in the reaction mechanism of the enzyme) is an important component in the assembly of Tgl. The recruitment of Tgl to the surface of the developing spore by its substrates directs the enzyme to a pre-formed protein assembly resulting in its fortification through covalent protein-protein cross-linking.

RESUMO

O processo de esporulação em *Bacillus subtilis* culmina com a formação de uma célula dormente denominada de endosporo ou esporo. Dos tipos de células conhecidas, o esporo é das mais resistentes e consegue permanecer viável durante longos períodos de tempo. A camada mais externa dos endosporos de *B. subtilis* é denominada de manto e contribui em larga escala para a resistência do esporo a insultos externos, sendo também importante para a monitorização e interacção do esporo com o ambiente que o circunda. O manto é composto por mais de 70 proteínas que são produzidas em diferentes etapas da esporulação e que se organizam à volta do esporo em desenvolvimento (pré-esporo), levando assim à construção de uma estrutura macromolecular. Um grupo de proteínas chave, denominadas morfogenéticas, são essenciais para a formação do manto. A ausência destas proteínas gera anomalias severas na organização/construção do manto, conduzindo por sua vez a defeitos funcionais. No entanto, muitos outros factores parecem contribuir para a correcta construção do manto, incluindo a formação de ligações covalentes entre as suas proteínas (*cross-linking* de proteínas).

Presente no manto encontra-se uma transglutaminase (TGase), a Tgl, que cataliza o *cross-linking* de pelo menos quatro proteínas do manto. As TGases são conhecidas pela sua capacidade de catalizarem a formação de ligações isopeptídicas do tipo ε-(γ-glutamil)lisil entre as cadeias laterais de resíduos de glutamina e lisina (substratos Q e K, respectivamente) presentes em proteínas, resultando no seu *cross-linking*. Estas enzimas estão frequentemente associadas à estabilização de tecidos ou estruturas macromoleculares tendo um papel importante em processos de desenvolvimento. A análise estrutural de TGases sugere que estas enzimas, juntamente com proteases de cisteína do género da papaína e endopeptidases de cisteína do peptidoglicano, evoluíram a partir de um ancestral estrutural comum. Apesar de as TGases estarem distribuídas pelos três grandes domínios celulares, o conhecimento actual destas enzimas é quase exclusivamente derivado de estudos com TGases animais. Até à data, apenas uma TGase bacteriana, de uma espécie de *Streptomyces*, foi analisada em detalhe,

mostrando que esta está longe do paradigma das TGases animais e sugerindo que ainda existe muito para aprender sobre as TGases bacterianas. Com o objectivo global de contribuir para o conhecimento estrutural e mecanístico das TGases bacterianas, decidimos analisar a Tgl. Determinámos a estrutura da enzima e analisámos a sua localização e função durante a esporulação, de forma a tentar correlacionar o papel e comportamento da Tgl *in vivo* com as suas características estruturais.

Começámos por resolver a estrutura da Tgl. Verificámos que, apesar da enzima apresentar uma estrutura global única, a conformação do seu núcleo catalítico aproxima-se da visualizada nas endopeptidases de cisteína do peptidoglicano, que por sua vez representam o actual parente mais próximo do ancestral estrutural comum das TGases e das proteases de cisteína do género da papaína. Estes resultados apoiam a contenção de que algumas TGases evoluíram a partir de um ancestral estrutural que é comum a diferentes famílias de enzimas. Ensaios enzimáticos com a forma nativa e mutantes pontuais da Tgl indicam que a actividade catalítica da enzima é realizada através de uma díada catalítica parcialmente redundante, composta por Cys116-Glu187/Glu115. Estes resultados contrastam com os relatados sobre TGases animais cuja actividade é dependente de uma tríada catalítica conservada, formada por Cys-His-Asp. É possível que a presença de uma díada catalítica seja uma característica prevalente em TGases bacterianas. A Tgl parece também ter uma estrutura simplificada quando comparada com a de outras TGases, e como tal, poderá representar os requisitos mínimos e necessários para a catálise de *cross-linking* de proteínas. Os nossos estudos estruturais e bioquímicos indicam que a Tgl é sintetizada numa forma activa em completa oposição ao que é visto nas outras TGases analisadas até à data. Uma característica notável da estrutura da Tgl é a presença de um túnel hidrofóbico que atravessa a molécula de um lado ao outro e em cujo centro está localizado o centro activo. Apesar de túneis semelhantes terem sido detectados em TGases animais, estes só se formam no estado activo das enzimas.

De seguida examinámos o papel da Tgl na formação do esporo. Surpreendentemente, os nossos resultados indicam que a enzima não está apenas associada com o manto mas também com o córtex dos esporos. O

córtex é uma camada interna e justaposta ao manto, constituída por uma forma modificada de peptidoglicano. SafA e a sua forma mais pequena, C30, duas proteínas do manto que são substratos da Tgl, estão também presentes no manto e no córtex, e a actividade da Tgl afecta ambas as proteínas em ambas as camadas do esporo. Como tal, os resultados indicam que o papel da enzima poderá ser o de cimentar a interface entre o córtex e o manto, contribuindo desse modo para a resistência final do esporo. SafA é uma das proteínas morfogenéticas cuja ausência durante a esporulação impede o recrutamento e correcta localização de diversas proteínas do manto. Os nossos resultados indicam que a eliminação de *safA* conduz a uma redução dos níveis da Tgl no manto e elimina a detecção da enzima no córtex. Assim, o estado final de SafA e Tgl no esporo, parece ser modulado por um circuito interdependente e auto-regulatório. Adicionalmente, ensaios de localização efectuados com uma fusão funcional Tgl-CFP mostraram que *safA* é importante para a localização correcta de Tgl-CFP durante o processo de esporulação, embora uma fracção significativa da enzima ainda apresente a localização correcta na ausência de SafA. Estes resultados indicam que existe redundância no processo de localização da Tgl para o pré-esporo.

Visto que SafA é um dos substratos da Tgl, investigámos se outros substratos da enzima seriam importantes para a sua localização no pré-esporo. Verificámos que a ausência dos substratos conhecidos da Tgl afecta o recrutamento da Tgl-CFP para o pré-esporo, embora os efeitos variem conforme os substratos. Adicionalmente, a deslocalização de C30, um dos substratos da Tgl, conduz à deslocalização de Tgl-CFP de forma idêntica. Finalmente, identificámos resíduos não catalíticos de Tgl, que estão presumivelmente envolvidos em interacções com substratos. A substituição individual destes resíduos para Ala conduz a defeitos no recrutamento da Tgl-CFP para o pré-esporo, sugerindo que o recrutamento e localização da enzima é dependente da localização dos seus substratos. O túnel detectado na estrutura da Tgl segrega os lados de interacção com os substratos Q e K, que estão localizados em entradas opostas do túnel. Identificámos os lados Q e K da Tgl e verificámos que os resíduos localizados em ambas as entradas do túnel, assim como a actividade catalítica da Tgl são importantes para o

recrutamento da enzima para o pré-esporo durante a esporulação. No entanto, apenas a interacção com o substrato Q aparenta ser de suma importância para o recrutamento da enzima. Esta observação em particular, sugere que a formação de um complexo covalente Tgl-substrato Q (que corresponde ao primeiro passo do mechanismo enzimático das TGases) é um importante componente no processo de recrutamento da enzima para o pré-esporo. O recrutamento da Tgl para a superfície do pré-esporo guiado pelos seus substratos, direcciona a enzima para uma estrutura proteíca pré-formada, resultanto em seguida na sua fortificação através da formação de ligações covalentes entre proteínas.

Chapter 1

General introduction

BACTERIAL ENDOSPORES

The ordered assembly of protein-based complexes can lead to the formation of intricate macromolecular structures, such as the eggshell, the flagella or the related type 3 secretion system, virus particles and the type IV pilus (1-4). Understanding how these structures are assembled has been the focus of investigation in different areas and one prime example is that of the formation of the coat layer of bacterial endospores (spores for simplicity). Spores are the end result of the sporulation process in species of *Bacillus*, *Clostridium,* and related organisms and are likely the most resistant cell type known, able to survive for hundreds of years, if not more (5). These metabolic dormant cells are able to withstand extremes of heat, chemical and physical stresses and UV radiation (5-7), just to germinate when conditions are favorable and commence growth and replication cycles. The coat not only contributes to the spore's overall resistance but it's also important for spore germination (6,8,9). In part due to their remarkable endurance properties, spores have been used in many different applications. *B. subtilis* (non-pathogenic) spores are in fact currently used as probiotics as they can be easily produced and formulated, they survive the harsh gut conditions and also have immune modulatory properties (10). In addition, the ability to display peptides or proteins (antigens or enzymes of commercial interest) at the spore surface, as fusions to coat proteins, has allowed a number of applications of spores in biomedicine (e.g., vaccinology) or biotechnology (11-13). As such, understanding how the coat is shaped is of undisputed importance. However, and despite several decades of study, some aspects of its construction still remain obscured. Several reasons may contribute to this, including its complex composition (over 70 proteins in *B. subtilis*), the absence of obvious phenotypes, under laboratory conditions, associated with disruption of many of the genes for coat components, the complex regulation of coat gene expression, and the range of post-transcriptional and post-translational mechanisms that contribute to the assembly of coat proteins (7,14-18).

With this work we aimed to contribute to a better understanding of the assembly processes of the spore coat. We focused on the role of Tgl, a post-

translational modification enzyme, in proper coat assembly. Tgl is a transglutaminase and this group of protein cross-linking enzymes is widespread in nature (19). Animal transglutaminases have been extensively studied and several structures are available which indicate that the enzymes all share a similar fold. However, the sequence of Tgl shows that the enzyme is only conserved in *Bacillus* and *Clostridium* species, which may reflect a unique enzyme with a novel fold, deviating from the animal paradigm. Furthermore, the current knowledge on bacterial transglutaminases is scarce. Thus, by studying Tgl, we also hoped to bridge the knowledge gap between animal and bacterial transglutaminases.

The following sections will summarize the relevant aspects of the sporulation process and coat assembly in *B. subtilis*. They will also focus on what is currently known about the transglutaminase group of enzymes.

SPORULATION IN *BACILLUS SUBTILIS*

Structure and functional characteristics of the spore. The spore is composed by three main concentric layers easily visualized by electron microscopy: the core, the cortex and the coat, with the core and coat as the innermost and outermost layers, respectively (Figure 1.1A).

The core contains a copy of the genome and presents a very low water content. The dehydrated state of the core is thought to be the major contributor to the spore's dormancy state. Inside the core there is also RNA, enzymes, ribosomes and high levels of the calcium chelate of dipicolinic acid (5,20,21). The core of the spore also contains unique small acid soluble proteins which bind directly to DNA, protecting it from damage. The characteristics and components within the core contribute to the spore's overall resistance to heat, chemicals and UV radiation (5,20,21).

Surrounding the core is the cortex layer. The cortex is composed by peptidoglycan, which differs from the cell wall peptidoglycan of growing *B. subtilis* cells. Peptidoglycan is a polymer, which consists of alternating *N*-acetyl-glucosamine (NAG) and *N*-acetyl-muramic acid (NAM) sugar residues, arranged as glycan strands. NAM residues carry pentapeptide side chains

which can be used in the cross-linking of different glycan chains. Due to the specific characteristics of the cortex peptidoglycan, this polymer is much less cross-linked than that of *B. subtilis* cells. ~50% of the NAM residues of the cortex peptidoglycan do not present amino acid side chains and are cyclized in the form of muramic-δ-lactam. Also, another 25% of the NAM residues only have L-alanine as a side chain. Thus, only 3% of the cortex NAM residues present side chains that are involved in cross-linking (in opposition to the 40% detected in the peptidoglycan of *B. subtilis* cells) (21,22). The low level of cross-linking in the cortex peptidoglycan is important for the maintenance and/or achievement of core dehydration and thus contributes to the spore's heat resistance. The cortex also possibly adds to the spore's resistance against certain chemicals (21,23,24).

The coat, the outermost layer of the spore in *B. subtilis*, is composed by four sub-structures: the basement, inner and outer coat layers, and the crust (from the cortex to the outside, Figure 1.1B) (25,26). However, by electron microscopy and conventional staining only two are easily distinguished: the inner (lamellar) and outer (striated, and more electron dense) coat layers (Figure 1.1A). The coat is mainly composed of protein (with more than 70 different polypeptides) but it also contains carbohydrate components (21). The proper assembly of the coat contributes to the spore's resistance to extremes of wet heat, UV radiation, lytic enzymes, noxious chemicals and mechanical stress and is also important for germination (6,7,21,27).

The spore also presents three other structures which are not always easily visualized by electron microscopy: the inner and outer forespore membranes, and the germ cell wall (Figure 1.1B). The inner forespore membrane encircles the core and is present in a semisolid state, with the lipids that compose it mostly immobile; this may be due to cortex compression or to the core's dehydrated state (21). The inner forespore membrane is most likely responsible for the very low permeability of dormant spores even to small molecules; also, it contains several proteins which are important for the proper germination of spores (20,21). The germ cell wall, composed by peptidoglycan of similar composition of growing *B. subtilis* cells, is assembled on the outside of the inner forespore membrane and it

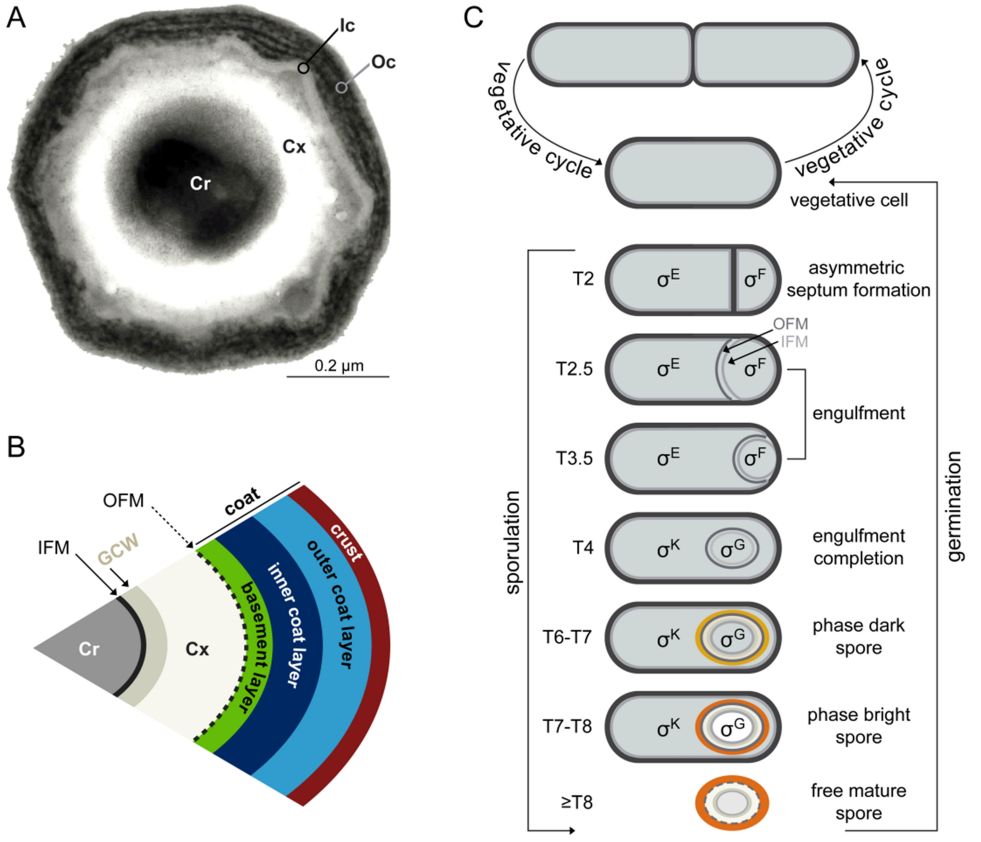


Figure 1.1. Structure of the spore and the sporulation process. A. Thin-section electron micrograph of a spore of *B. subtilis*. **B.** Diagram of all the different structures of the mature spore. Note that the outer forespore membrane (OFM) is represented by a dash line as it is not known if the membrane remains intact in mature spores. **C.** During growth, in the vegetative cycle, *B. subtilis* divides symmetrically to create two cells of identical size. The sporulation process, which takes about 8 hr to complete, is regulated by four sigma factors of RNA polymerase and can be followed through a series of morphological stages. First, an asymmetric septum is formed near one of the poles of the vegetative cell, leading to the formation of two distinct compartments: the mother cell (bigger) and the forespore (smaller); septum formation triggers the activation of σ^F in the forespore which in turn activates σ^E in the mother cell. The peptidoglycan within the septum is hydrolyzed and the two septal membranes migrate around the forespore in a process denoted as engulfment. Once the engulfment process is complete, σ^G becomes active in the forespore which leads to the activation of σ^K in the mother cell. The germ cell wall is then synthesized and, soon after, cortex synthesis begins, leading to the appearance of phase dark spores inside the mother cell. With cortex maturation phase bright forespores become visible. Finally, the mother cell lyses and the spore is released. Coat proteins are synthesized under the control of σ^E and σ^K, and coat assembly starts just after asymmetric division (T2.5), continuing until mother cell lysis; complete spore maturation is only achieved after mother cell lysis. The mature free spore can then germinate when conditions are favorable and a new vegetative cycle can begin. The coat is represented in orange. The numbers on the left hand side correspond to the time in hr after sporulation initiation. Cr, core; Cx, cortex; Ic, inner coat; Oc, outer coat; GCW, germ cell wall; IFM, inner forespore membrane; OFM, outer forespore membrane.

will be the cell wall of the new cell generated by germination (22). Unlike cortex peptidoglycan, the germ cell wall peptidoglycan does not contain muramic-δ-lactam. This difference allows for the specific degradation of cortex peptidoglycan, but not of the germ cell wall during germination, as the cortex lytic enzymes specifically recognize a substrate containing muramic-δ-lactam (22,28). Finally, the outer forespore membrane is located between the cortex and coat layers (Figure 1.1B). It is not known if this membrane is present, or if remains intact, in mature dormant spores (21).

Morphological stages coupled to gene expression regulation. During growth (vegetative cycle), *B. subtilis* divides symmetrically to give rise to two equal sized daughter cells (Figure 1.1C). However, in conditions of nutrient depletion, it has the ability to initiate sporulation, a developmental process which takes about 8 hours to be completed under standard laboratory conditions (Figure 1.1C). The first blatant morphological stage in sporulation is the formation of a division septum near one of the cell poles (29). This asymmetric division creates two independent but interconnected compartments of different sizes: the mother cell (larger) and the forespore (smaller). In the end, the forespore will develop into a mature spore at the expense of the mother cell, which will lyse. Soon after asymmetric septum formation, the two membranes start to migrate around the forespore in a process known as engulfment (29,30). Engulfment completion creates a free protoplast inside the mother cell surrounded by two membranes of different polarity (inner and outer forespore membranes). Engulfment completion triggers the last stages of sporulation and the germ cell wall and cortex layers start being synthesized (31), leading to the appearance of phase dark spores inside the mother cell, as detected by phase contrast microscopy. Soon after, phase bright forespores can be visualized, coinciding with cortex maturation processes, and finally the sporulation process ends with mother cell lysis and spore release. However, complete spore maturation is only achieved after mother cell lysis; for instance, full spore heat resistance as well as certain post-translational modifications of spore coat proteins are only detected following the spore's release from the mother cell (7,27,32).

The time and place of expression of sporulation genes is rigidly regulated and is under the control of four specific sigma factors of RNA polymerase: σ^F, σ^E, σ^G and σ^K (Figure 1.1C). Once the asymmetric septum is formed, the two cells become physically separated and they have two distinct, although interconnected, gene expression programs. The activation of the different σ factors is coupled to specific morphological stages of sporulation (reviewed in (30,33)). Asymmetric septum completion allows for the activation of σ^F in the forespore which in turn triggers the activation of σ^E in the mother cell. After engulfment completion, σ^G is active in the forespore and triggers σ^K activation in the mother cell. Five ancillary transcription regulators are also present during sporulation. Two of them, SpoIIID and GerE, have central roles in the control of the expression of several sporulation genes in the mother cell, either by repression or activation. SpoIIID is part of the set of genes that is transcribed under the control of σ^E, while transcription of *gerE* is dependent on σ^K. The coat proteins are synthesized in the mother cell compartment under the control of, either σ^E or σ^K, and deposited on the surface of the developing forespore. On the other hand, germ cell wall and cortex formation, whose significant synthesis is only detected after engulfment completion, are dependent on genes under the control of all four sporulation specific σ factors (31,34,35).

Spore coat formation. *The spatial arrangement of the coat layers is regulated by morphogenetic proteins*. The spore coat proteins are synthesized in the mother cell compartment and deposited around the developing forespore (7,14-16,25). Although it's not completely understood how all the different proteins are orderly assembled, a small subset, called morphogenetic proteins, has a major role in spore coat formation. The formation of each coat layer is dependent on its own morphogenetic protein (Figure 1.2). SpoIVA guides the formation of the basement layer. In its absence no cortex is formed and the coat accumulates as swirls of material in the mother cell cytoplasm (36,37). Nonetheless, this coat material has the appearance of an assembled coat which is simply not attached to the forespore. Thus, the role of SpoIVA appears to be that of marking the outer forespore membrane as the site for spore coat deposition. SafA and CotE are responsible for inner and outer coat formation,

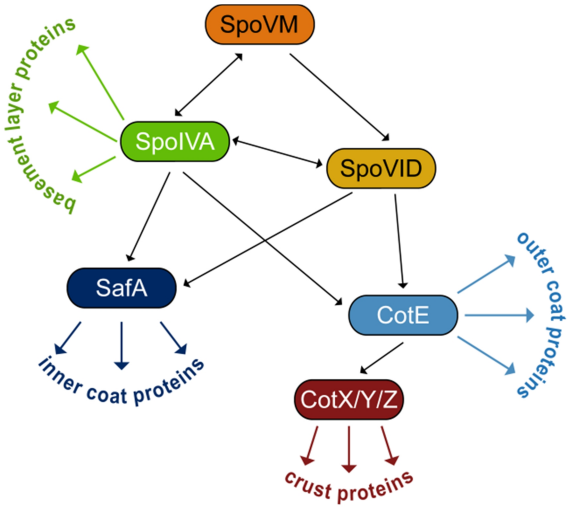


Figure 1.2. Spore coat formation is controlled by a hierarchy of assembly dependencies. The formation of each coat layer is under the control of its own morphogenetic protein: SpoIVA, SafA, CotE and CotX/Y/Z. SpoVM and SpoVID are responsible for the encasement process (see text for details). The diagram represents the assembly dependencies between different coat proteins and highlights the central role of morphogenetic proteins. Note however that the representation is simplistic as the role of each morphogenetic protein is not limited to the formation of its layer.

respectively (26,38-40), and finally, CotX, CotY and CotZ are determinant for crust formation (26,41) (Figure 1.2). These morphogenetic proteins guide spore coat assembly and create a scaffold to which coat proteins can be attached.

Two other proteins have a critical role in spore coat assembly: SpoVID and SpoVM (Figure 1.2). The concerted action of these two proteins is important for the localization of spore coat proteins as a shell around the forespore, a process known as encasement (42). Although both SpoVM and SpoVID are necessary for proper encasement, in the absence of SpoVID, SpoVM is still able to form a shell around the forespore, while the same is not true for SpoVID. Thus, SpoVM sits on top of the hierarchy that controls spore encasement (42) (Figure 1.2). SpoVM also has extra roles during sporulation (see below) and in its absence an ill-defined cortex and a thin, loosely attached coat are detected (43). In the absence of SpoVID, spore coat proteins fail to encase the spore and accumulate as swirls of material in the mother cell cytoplasm, a phenotype reminiscent of a *spoIVA* mutant. However, the absence of SpoVID does not impair cortex formation (44).

There appears to be a hierarchy in coat assembly. The formation of the basement layer is required for the formation of all other layers, and the crust is dependent on outer coat assembly. However, in the absence of the inner coat, the outer coat proteins are still assembled and vice-versa. Furthermore, some proteins synthesized early in sporulation are present in

the outermost layers of the coat, while some of the constituents of the basement and inner layers are expressed late. This implies that the assembly of the coat is not done by the ordered addition of each layer (from the inside to the outside), but rather, all the layers are simultaneously assembled (14,26,45). In fact, all the morphogenetic proteins described above are expressed early in sporulation under the control of σ^E (with the expression of CotE and crust morphogenetic proteins additionally regulated by σ^K) and their localization is detected as soon as the asymmetric septum starts to curve (45). This initial localization of the morphogenetic proteins is already spatially organized into the four distinct coat layers that are detected in phase bright forespores (Figure 1.3) (26,45).

Encasement regulates spore coat assembly. A recent study indicates that spore coat assembly is regulated by multiple waves of encasement which are coupled to the transcriptional regulation of the genes for the various coat proteins (45). Proteins synthesized prior to engulfment are detected on the mother cell proximal pole of the forespore (Figure 1.3). Once engulfment is complete, spore coat proteins tend to localize at both poles of the forespore (Figure 1.3). Encasement of the different spore coat layers is visualized at different times. SpoVM and SpoVID are the first proteins to encircle the spore, tracking along with the engulfing membrane, and encasement progresses from the basement layer (concurrent with the engulfment process), followed by the inner and outer coat layers (after engulfment), and finally the crust (with the appearance of phase bright forespores) (Figure 1.3) (45). While the encasement of the inner coat is delayed in relation to that of the basement layer, some inner coat proteins (probably including SafA) encase the spore as engulfment proceeds (45).

The temporal differences in the encasement of the distinct coat layers reflect the time and transcriptional dependencies of coat genes. The basement and inner coat layers depend on SpoIVA and SafA, respectively, and their expression is under the control of σ^E, although expression of *spoIVA* is likely repressed by SpoIIID (38,45,46). *cotE* transcription takes place from two promoters, P1 and P2 (47,48); *cotE* expression from the P1 promoter is under the control of σ^E and is repressed by SpoIIID, while transcription from P2 is regulated by σ^K and repressed by GerE (46). Finally,

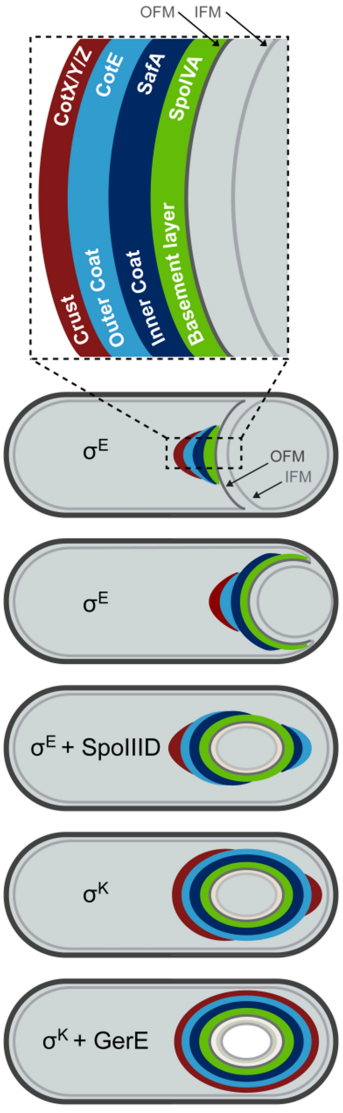


Figure 1.3. Spore coat assembly is regulated by the spatial arrangement of the different coat layers and by successive waves of encasement. Coat assembly starts as soon as the asymmetric septum starts to curve. At that time, the morphogenetic proteins that control the assembly of each of the coat layers are detected in the mother cell proximal pole (MCP) of the forespore and spatially define the place of assembly of each layer (as illustrated in the zoomed view at the top). Successive waves of encasement will regulate the assembly of the spore coat; the encasement of the basement layer proteins appears to be concurrent with the engulfment process; the inner and outer coat layers are slightly delayed and just after engulfment completion a second cap is detected on the mother cell distal (MCD) pole; later the two caps will be connected to form a complete shell around the forespore (nonetheless, the encasement of some inner coat proteins, such as SafA, seems to occur sooner, concomitantly with engulfment); finally, the crust proteins are only detected in the MCD pole of the forespore once phase dark spores are visualized. The temporal variances of the different waves of encasement are related to the transcriptional control of the different coat proteins, as multimerization of the morphogenetic proteins, that control the assembly of each coat layer, is likely important for encasement completion. IFM, inner forespore membrane; OFM, outer forespore membrane.

transcription of the genes for the crust morphogenetic proteins is activated by σ^E, σ^K, and GerE (46,49). As such, pulses of gene expression regulate the encasement process (Figure 1.3). It has been shown that SpoIVA, SpoVID, SafA, CotE, and CotY/CotZ form multimers ((50-52,53,54,55), and our unpublished data, Pais, T.). Thus, multimerization of the morphogenetic proteins may play a role in the formation of the scaffold with which other proteins can interact.

SpoIIQ, a forespore protein shown to be important for encasement (45), localizes to the inner forespore membrane and interacts with SpoIIIAH (a

mother cell protein) which localizes to the outer forespore membrane (56,57). The SpoIIQ-SpoIIIAH complex is responsible for the creation of a channel that spans across the inner and outer forespore membranes allowing for the connection of the mother cell and forespore compartments (58,59). This channel contributes to the engulfment process and is essential for the activity of σ^G following engulfment completion (57,60,61). In the absence of SpoIIQ, the encasement process is severely compromised, much more than when SpoIIIAH is absent or when *sigG* is eliminated (45). Although the role of SpoIIQ in encasement is not completely understood, this observation indicates that forespore proteins also contribute to coat assembly.

Interactions between morphogenetic proteins. Another driving force behind the encasement process and overall spore coat assembly is protein-protein interactions among morphogenetic proteins (Figure 1.4A). The localization of SpoVM and SpoIVA is mutually dependent. SpoVM, a small polypeptide of 26 amino acids, recognizes the positive curvature of the outer forespore engulfing membrane and is oriented parallel to it (62,63). SpoVM interacts with SpoIVA and is responsible for its proper localization and anchoring to the outer forespore membrane. On the other hand, the proper localization of SpoVM is dependent on SpoIVA, and in its absence, SpoVM fails to localize exclusively to the outer forespore membrane (62). SpoIVA localization is also dependent on its Walker A nucleotide-binding motif (Figure 1.4B) which binds and hydrolyzes ATP (50). ATP hydrolysis appears to drive the self-assembly and multimerization of SpoIVA which is important for its proper localization during sporulation (but not for the localization of SpoVM). Disruption of the Walker A motif impairs the formation of the coat structure. Thus, SpoIVA multimerization, induced by ATP hydrolysis, is important for the formation of the basement layer, creating a platform for the addition of the subsequent coat layers (50).

SpoIVA then interacts with SpoVID, recruiting it to the forespore (Figure 1.4A) (42,54). Region A, in the C-terminal domain of SpoVID (Figure 1.4B), is most likely involved in the direct interaction with SpoIVA, and is important for the targeting and localization of SpoVID to the forespore surface. The proper localization of SpoVID is also dependent on its LysM domain

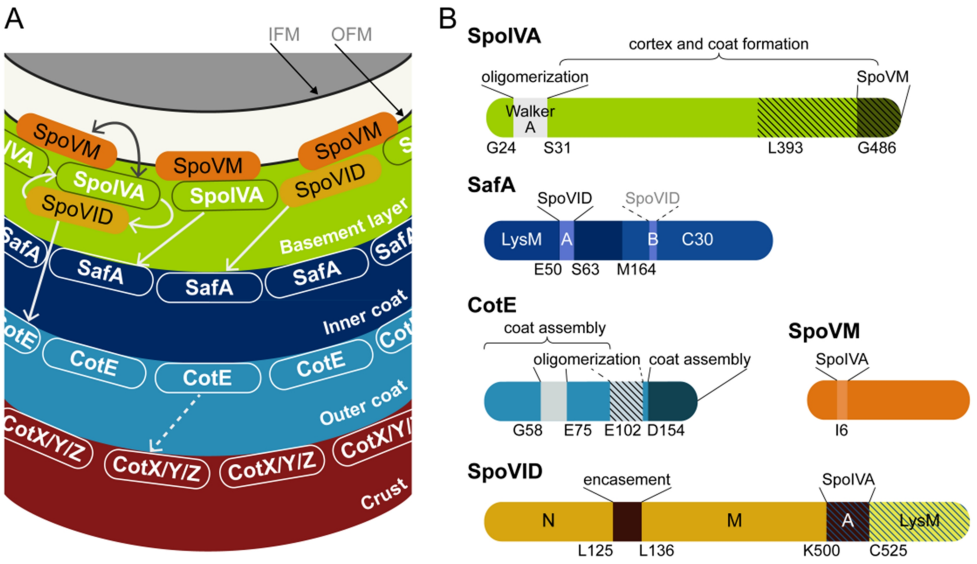


Figure 1.4. Interactions between morphogenetic proteins drive spore coat assembly. A. Representation of the interactions formed among morphogenetic proteins during spore coat formation. SpoVM localizes to the outer forespore membrane and anchors SpoIVA; on the other hand, SpoIVA is necessary to restrict the localization of SpoVM solely to the outer forespore membrane. SpoIVA interacts with SpoVID and SafA and is likely responsible for the recruitment of both proteins to the forespore. Additionally, it has been seen that SpoVID interacts with both CotE and SafA; it is likely that these interactions, along with the SpoVID-SpoIVA interaction, are important for the encasement process. It is also probable that CotE will directly interact with one or more of the crust morphogenetic proteins even though, such an interaction has never been reported (dashed arrow). Proteins are not drawn to scale; IFM, inner forespore membrane; OFM, outer forespore membrane. **B.** Schematic representation of the primary structure of the major morphogenetic proteins with structural and functional domains highlighted (crust morphogenetic proteins are not represented as there is currently no data). The striated pattern represents regions important for the proper localization during sporulation. Note that in SpoIVA and CotE, regions spread throughout the sequence are necessary for spore coat assembly, unlike SpoVID where only a small region of the protein appears to be necessary for encasement. With the exception of SpoVM, proteins are represented with an approximate scale. Panel B was created based on data from (24,42,50,51,54, 62,64-67).

(peptidoglycan binding domain) present at the C-terminus (Figure 1.4B) (42). In the N-terminal part of SpoVID a stretch of 12 residues is essential for encasement of proteins from all the different coat layers (64) (Figure 1.4B), including SpoIVA. This implies that a second interaction between SpoIVA and SpoVID most likely occurs during encasement. Finally, because the encasement of SpoVID is dependent on SpoVM (42), it is also expected that the two proteins directly interact during the process, although this interaction has never been demonstrated.

The next morphogenetic protein in the organization of spore coat

assembly is SafA. SafA is able to directly interact with SpoVID and SpoIVA, and the three proteins are able to form a tripartite complex, although the SpoIVA-SafA interaction appears weaker than that of SpoVID-SafA (52,54,64,66,68,69). Region A (composed of 13 residues), in the N-terminal part of SafA (Figure 1.4B), is necessary for the SafA-SpoVID interaction (66). A second site, named region B (Figure 1.4B), also appears to contribute, to a lower degree, to the interaction of the two proteins. While region A alone is sufficient to promote the SpoVID-SafA interaction, the same is not true for region B (66). Like SpoVID, SafA also contains a LysM domain (Figure 1.4B); however its role is unclear, as its elimination appears to generate a form of SafA which fails to accumulate during sporulation (66).

CotE and SpoVID strongly interact *in vitro* (55,64). CotE is initially targeted to the forespore in a SpoVID-independent manner; immuno-electron microscopy shows that CotE localizes as a ring at a distance of ~75 nm from the forespore membrane. This distance is maintained during engulfment and most likely corresponds to the site of deposition of basement and inner coat proteins (39). It is not clear how CotE and SpoVID interact *in vivo*, seeing that the two proteins are present in different coat layers, which are not juxtaposed (Figure 1.4A). One possible explanation is that only a few molecules of SpoVID will be needed to drive CotE encasement, and as such, SpoVID localization outside of the basement layer is not detected. This view is credited not only by the strong SpoVID-CotE interaction, but also by the ability of CotE to multimerize (47,51,55). Furthermore, the M domain of SpoVID (Figure 1.4B) appears to be intrinsically disordered as predicted by PSIPRED (70) and PONDR (71-73) servers, unlike the N and C-terminal parts (which control encasement and SpoVID localization, respectively). Thus, SpoVID may be attached to the basement layer by its C-terminal region but extends beyond the inner coat layer to interact with CotE, through its N-terminal domain. Although sequences responsible for localization, oligomerization and, outer coat assembly functions have been mapped in CotE (Figure 1.4B) (51,65), the residues involved in the interaction with SpoVID are still unknown.

Finally, because CotZ localization is dependent on CotE (14,45), it is likely that CotE interacts with one or more of the morphogenetic crust proteins.

Nonetheless, a direct interaction between CotE and any of the crust proteins has never been shown although, evidence suggests that CotE and CotY establish a low-affinity interaction (53).

Additional factors involved in proper coat formation. While the morphogenetic proteins have decisive roles in spore coat assembly there are other factors that influence proper coat construction.

Cortex and coat formation appear to be mainly independent and mutants have been isolated that display only one or the other (36,44,74). However, in a *spoIVA* mutant both cortex and coat formation are abolished, implying that the initiation of cortex and coat formation is coupled. In fact, a recent study identified a small protein of 37 residues, denoted CmpA, which along with SpoVM, creates a developmental checkpoint that delays cortex synthesis until coat assembly is initiated (75). Overexpression of CmpA abolishes cortex synthesis, while *cmpA* elimination leads to the appearance of heat resistant spores (indicative of cortex maturation) sooner than in the wild type strain, suggesting that the sporulation process occurs faster in the absence of CmpA (75). Interestingly, while the elimination of *cmpA* does not seem to cause drastic changes in the overall spore structure as detected by electron microscopy, the spores are less resistant to lysozyme. This is a clear indication that coat formation was abnormal, and most likely the defects of the *cmpA* mutant spores are related to spore maturation processes that occur late in sporulation, as the initiation of spore coat assembly and CotE localization appears normal in the mutant (75). Thus, proper coat formation also relies on proper cortex formation.

While the formation of each layer of the coat depends on a specific morphogenetic protein, the role of the major morphogenetic proteins is not limited to the formation of its layer, rather all layers appear to be interconnected and interdependent. For example, CotS, a protein that localizes to the inner coat and at the edge of the cortex, depends on CotE for assembly (76). Another example is that of OxdD, whose localization is dependent on SafA and independent of CotE, but fails to remain associated with free spores in the absence of CotE (77). The assembly of CotG, a protein of the outer coat, depends on both SafA and CotE (38,47).

Furthermore, the major morphogenetic proteins alone do not control the assembly of individual coat proteins. In effect, there appears to be an hierarchy in protein interactions and a prime example is that of the CotE-dependent assembly of CotH, another morphogenetic protein (78). Spores of a *cotH* mutant show a disordered outer coat as visualized by electron microscopy (78-80). However, CotH also has a CotE-independent role and is involved in proper inner coat formation (78,81). In fact, CotH appears to be associated with the inner coat (78,80). CotH is necessary for the localization of CotQ, CotU and CotZ (among others) (14); noteworthy is the fact that CotZ is a crust protein (25,26). Additionally, CotH affects the assembly of CotG and CotS (14,79,80). Indirectly this affects the assembly of CotB and CotSA which are dependent on CotG and CotS, respectively (82,83). There are also examples of isolated assembly dependencies which additionally contribute for the overall formation of the spore coat. The localization of CotJA and CotJC is interdependent and the two proteins are thought to form a complex before localizing to the forespore (84). Finally, GerQ, an inner coat protein (85), is necessary for the localization of CwlJ, a cortex-lytic enzyme (86).

Final maturation processes are also important for the proper formation of the spore coat. During the assembly of the spore coat, the outer layers remain permeable, as it has been seen that some basement and inner coat proteins (with a maximum size of 70 kDa) are expressed late under the control of σ^K (45). Free mature spores, however, are resistant to lysozyme (~14 kDa) indicating that coat permeability changes in the very late stages of sporulation. Protein cross-linking events also occur late in sporulation, after mother cell lysis (see next section). Additionally, the spores' full wet heat resistance is only achieved hours after its release and this is linked to proper coat (and possibly, cortex) formation (27).

Also contributing to the intricate process of coat formation are post-translational modifications of coat proteins, including glycosylation, proteolysis (17) and phosphorylation (Freitas, C., Plannic, J., Serrano, M. and Henriques, A., unpublished results from our laboratory and (87)). The identity of the enzymes involved in the post-translational modifications is mostly unknown. However, one protein, YabG, which displays *in vitro*

protease activity, has been implicated in the proteolytic processing of several coat proteins (88,89). When the spore is fully formed about 30% of the spore coat material resists solubilization by different methods (90). YabG contributes to this highly cross-linked material, as in its absence several proteins are considerably more extractable than in wild type spores. Three types of covalent protein cross-linking have also been proposed or detected in this insoluble spore coat material: di-sulfide, di-tyrosine and ε-(γ-glutamyl)lysyl bonds (18,91,92). Tgl, a transglutaminase, catalyzes the later type of protein cross-linking contributing to the creation of the coat insoluble material.

***Tgl*.** Tgl appears to be conserved in only some species of the *Bacillus* and *Clostridium* genera indicating that the enzyme may have evolved to work specifically in the sporulation process. Its expression is activated by σ^K and repressed by GerE (46,93-96). Tgl-GFP localization during sporulation has been reported as strongly dependent on CotE (14,45,95). The assembly of the enzyme is also dependent on SafA and GerE (in the absence of this transcription factor the spores do not have an inner coat and display a defective outer coat (97)). The assembly dependence of Tgl on all these different factors indicates that the enzyme is likely present within different layers of the coat lattice. Nonetheless, because the localization of Tgl-GFP during sporulation is more affected by the absence of CotE than either SafA or GerE, just like the levels of the enzyme in free spores, Tgl has been described as a predominantly outer coat protein (14,26).

Disruption of the *tgl* gene leads to small defects in outer coat morphology as visualized by electron microscopy (98). *tgl* mutant spores display a decrease in wet heat resistance similar to what is seen in double *cotE gerE* spores (which lack most of the coat layers and only a small rind of material is detected around the spore) (27,99). Some studies have additionally reported a slower germination rate for *tgl* mutant spores, although this result is not consistent between different laboratories which probably reflects differences in methodology (14,93,95,98). Fluorescence microscopy experiments indicate that the assembly of YutH-GFP, a coat protein of unknown function, is dependent on Tgl (14). Finally, the analysis

of the soluble coat proteins by SDS-PAGE reveals that in *tgl* mutant spores several proteins display an increased extractability when compared to wild type spores (14,95,98). These coat proteins with increased extractability are most likely direct substrates of Tgl.

Four substrates of Tgl have been proposed: YeeK, GerQ (*ywdl*) and two forms of the SafA protein (93,95,100). During sporulation three different forms of the SafA protein are formed: SafA, $SafA_{N21}$ (N21) and $SafA_{C30}$ (C30) (67), due to an alternative site of translation initiation that starts at Met164 (Figure 1.4B). SafA corresponds to the full length form of the protein, which has the morphogenetic role during sporulation, while N21 and C30 correspond to the N- and C-terminal segments of the protein, with apparent migration sizes by SDS-PAGE of 21 and 30 kDa, respectively (67,88). Tgl appears to be involved in the cross-linking of SafA and C30 but not of N21 (95). Interestingly, all the know substrates of Tgl have been reported as inner coat proteins (25,68), which appears in contradiction with the strong CotE-dependency of Tgl for assembly. However, as discussed in the previous section, CotE-dependency does not necessarily mean that a protein localizes to the outer coat.

The activity of transglutaminases is highly regulated (101) (see next section). Tgl, however, seems to be produced in active form and no cofactors for enzymatic activity have been identified (102). Yet, the activity of Tgl *in vivo* seems to be delayed and to follow mother cell lysis (93,98). It has been suggested that the activity of Tgl may be regulated by the processing of its substrates as there seems to be a functional relationship between YabG and Tgl (95). As described in the previous section, the absence of YabG leads to the increase extractability of several proteins, which include the proposed substrates of Tgl (95). With the exception of GerQ, all of the other substrates of Tgl appear in a *yabG* mutant as slower migration species in an SDS-PAGE gel, when compared to *tgl* mutant spores. This suggests that YabG processes the substrates of Tgl and then the enzyme catalyzes the cross-linking reaction (95). In the same study, the authors show that the formation of higher molecular weight forms of GerQ can be overcome in the *yabG* mutant if the purified spores are incubated at 60° C (95). At this temperature Tgl displays higher activity (94,103), and the

formation of the high molecular weight products of GerQ at 60° C, in the *yabG* mutant, was Tgl dependent (95). Although the results suggest that the activity of Tgl is regulated by changes in its substrates, the direct link between processing and cross-linking has not yet been proved.

In short, Tgl is involved in proper coat formation contributing to the highly insoluble coat material that is detected in free spores. Even though the enzyme is expressed from about hour 4 of sporulation (98) its activity is only detected in the final maturation stages of the spore, presumably after mother cell lysis (93,98).

TRANSGLUTAMINASES

Transglutaminases (TGases) are ubiquitous enzymes (19,104) involved in several developmental processes, such as the formation of hair or skin, or in clotting structures (101,105). TGases catalyze transamidation reactions between the side chain of glutaminyl residues (Q substrate) and primary amines (K substrates) which can vary from simple molecules, naturally occurring polyamines or peptide-bond lysil residues (101,106,107). Protein cross-linking results from the reaction between protein-bound Gln and Lys residues leading to the formation of ε-(γ-glutamyl)lysil isopeptide bonds. The *in vivo* covalent attachment of proteins to lipids in a TGase-dependent manner has also been proposed (108). Recent studies have also shown that animal TGases are responsible for deamination reactions (the donor Gln residue is exchanged to Glu) (109). In recent years, the transamidation reactions catalyzed by TGases have been exploited in applications ranging from the construction of matrices for regenerative medicine (110) to site-specific modification (111) or labeling of proteins (112).

Animal TGases. *Catalytic residues and overall fold*. Because the malfunction of human TGases has been implicated in several diseases (113), they have been the most intensively studied, and most of the available structural information also pertains to this group of enzymes or to animal homologues. Animal TGases show a high degree of structural similarity and identical domains, as exemplified by the structures of blood coagulation

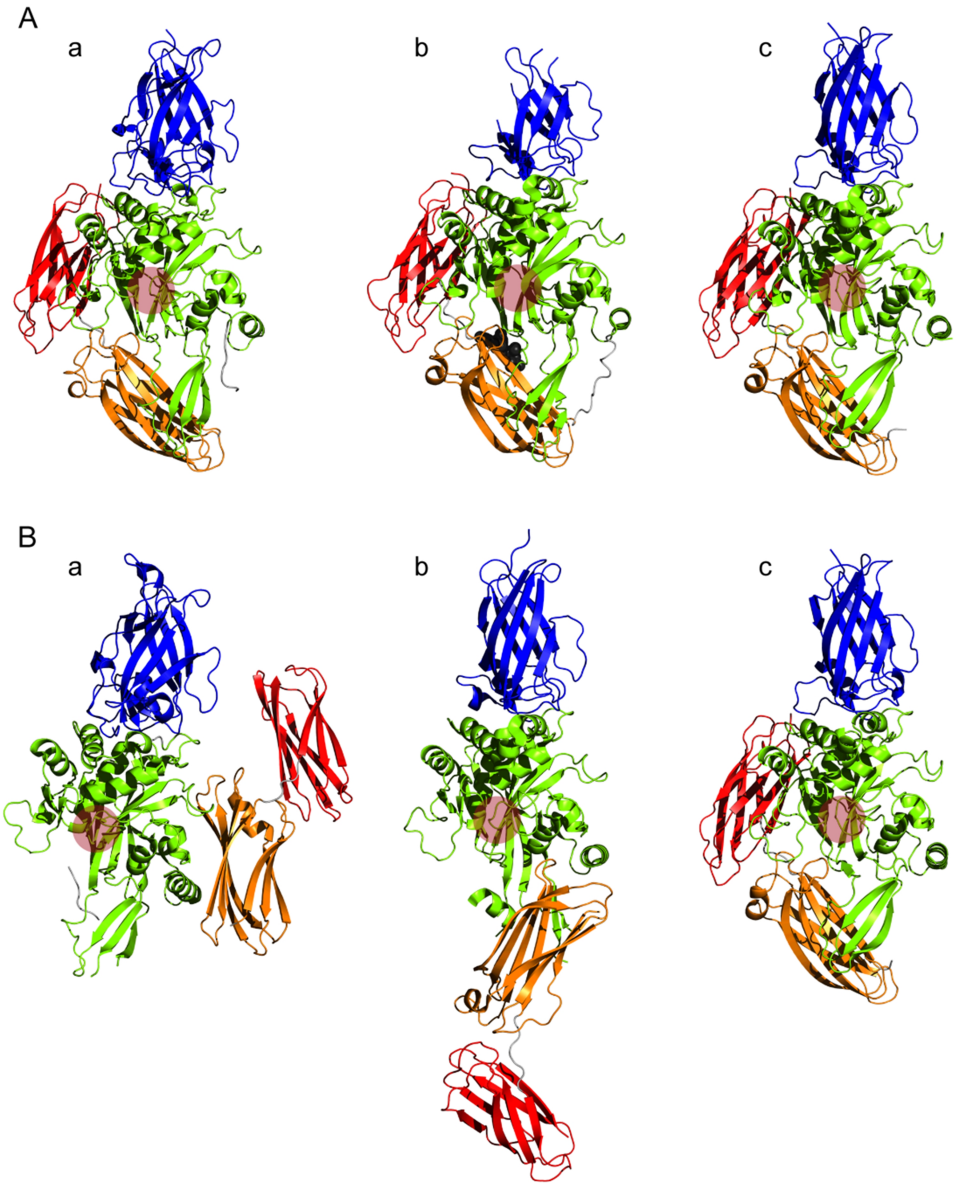


Figure 1.5. Cartoon representation of the inactive (A) and active structures (B) of Factor XIII/Factor XIIIa (a), TGase 2 (b) and TGase 3 (c). Structures are colored according to the different domains present in animal TGases: blue, β-sandwich; green, catalytic core; orange, β-barrel 1; red, β-barrel 2. The red circle represents the site of the catalytic triad. The inactive TGase 2 structure (pdb: 1KV3) corresponds to TGase 2 bound to GDP; GDP is shown as black spheres. The structures of Factor XIIIa (Ba) and active TGase 2 (Bb) correspond to the complex of the enzyme irreversibly bound to an inhibitor (the inhibitor is not shown for simplicity). The following pdb files were used: Factor XIII (Aa), 1GGU; inactive TGase 3 (Ac), 1L9M; Factor XIIIa (Ba), 4KTY; active TGase 2 (Bb), 2Q3Z; active TGase 3 (Bc), 1L9N.

Factor XIII (FXIII), TGase 2 and TGase 3 (Figure 1.5A) (106,107,114). All known TGases present an identical atomic arrangement in the catalytic core and are members of the papain-like superfamily of cysteine proteases (19,104,115,116). The members of this family possess a catalytic triad composed of Cys-His-Asp or Cys-His-Asn. Due to the specific arrangement of the catalytic core, it has been proposed that cysteine proteases and TGases might share a common evolutionary lineage derived from a minimal ancestral structural unit of the thiol-protease fold, which may correspond to the NlpC/P60 domain of cell wall endopeptidases (117). The NlpC/P60 domain with its simpler topology appears as a close relative of the minimal thiol fold (118) and these proteases also show a catalytic triad composed of Cys-His, and a third conserved polar residue (118,119). However, the NlpC/P60 fold is not easily recognizable in all the TGases structurally characterized until now.

Activity regulation. The activity of animal TGases is tightly regulated. They all require calcium to be activated and are produced and stored in zymogenic or otherwise inactive forms; in some cases their activity is also negatively modulated by GTP/GDP or ATP (101,109,114). The active form of animal TGases displays structural changes in the active site region when compared to the inactive enzyme (compare panels A and B, in Figure 1.5; the site of the catalytic residues is highlighted by a red circle). These structural changes are induced by the binding of calcium ions and lead to the exposure of the catalytic Cys residue, which in the inactive form of TGases is inaccessible from the outside (107,120-123). Also, the structures of active Factor XIIIa (FXIIIa) and active TGase 2, bound to an inhibitor that mimics a natural substrate (Figure 1.5Ba,b), reveal that the reaction with the substrate involves major conformational transitions in the structure of the enzymes (120,123). It is likely that the same is true for other animal-like TGases.

At least for TGase 2, the large complexity of the enzyme comes from its different functions, most of which are transamidation-independent. TGase 2 is present in most cell types and displays different subcellular localizations. Along with TGase 2 most cells express another active TGase. This has possibly facilitated the acquirement of different functions by TGase 2 which

appears to function as regulator (109,124). TGase 2 has been shown to function as a protein disulfide isomerase, a transcription regulator, to act as a scaffold protein and possibly also as a kinase and DNA hydrolase (109,124).

A very well documented activation mechanism of TGases is that of the emblematic plasma FXIIIa which is involved in the last stage of the blood coagulation cascade. The inactive form of the enzyme, FXIII, exists as an A_2B_2 heterotetramer, of which A_2 is the TGase zymogen, while B_2 is a non-enzymatic glycoprotein. Thrombin (another factor of the blood clotting process) cleaves the A subunits which weakens the interaction between the A and B subunits of the A_2B_2 complex. The dissociation of the complex however is only achieved by the binding of Ca^{2+}. The free A_2 subunit then undergoes conformational changes, in a calcium-dependent manner, leading to the unmasking of the active site cysteine and to the formation of the catalytic active, FXIIIa (125). FXIIIa will then cross-link fibrin and other proteins to create the stable and enduring blood clot. The activation of plasma FXIII is also regulated by its substrate, fibrin. Fibrin is generated by the processing of fibrinogen by thrombin. Fibrinogen processing exposes new end groups which allow for the self-assembly process of fibrin monomers into an end-to-end conformation of fibrin units (126). This end-to-end conformation is important for FXIIIa activity as the enzyme's activity is greatly enhanced when its cross-linking sites are present in linear assemblies (101,125). Thus the activity of FXIIIa is regulated by the assembly of its substrates. Furthermore, fibrin acts with thrombin to accelerate the processing and dissociation processes of FXIII into the A_2 + B_2 subunits. In fact, fibrin alone (even in the absence of calcium) can generate the dissociation of FXIII. This represents a feed-forward loop that is probably in place to make sure that FXIIIa is generated when fibrin is available but not fibrinogen. This is consistent with the fact that the majority of FXIII present in the plasma is bound to fibrinogen by its B_2 subunit. Although in place to catalyze the formation of a blood clot, only the presence of fibrin will lead to formation of active FXIIIa (125).

Curiously there is another example of substrate-dependent activity regulation of a TGase. TGase 1 is involved in the formation of the skin where

it contributes to the construction of the cornified envelope in the final stages of keratinocytes differentiation (107,127). The enzyme is synthesized in zymogenic form and is a membrane-bound TGase anchored by its N-terminal domain. TGase 1 is activated by calcium and by two proteolytic events at the end of the membrane-anchor and catalytic domains (107,127). Although some of the enzyme is located in the cytosol, the majority of the highly active TGase 1 detected in keratinocytes is present as a membrane-bound processed form, where the N-, catalytic and C-terminal domains remain in a complex which is maintained by secondary interactions (127). A recent study identified an additional factor, tazarotene-induced gene 3 (TIG3), which is able to activate TGase 1 (128). TIG3 is involved in several processes within the cell. It also associates with the membrane where it co-localizes with TGase 1 and this co-localization relates with the activation of TGase 1 activity. TIG3 appears to interact with TGase 1 (either directly or in a complex) which leads to the activation of the enzyme. On the other hand TIG3 itself serves as a substrate of TGase 1.

Non-animal TGases. Unlike animal TGases, reports on unrelated TGases are sparse and only two non-animal TGases structures are reported in the Protein Data Bank: that of GP42 from the oomycete *Phytophthora sojae*, and bacterial MTG from *Streptomyces mobaraense* (129-131). Both structures show an overall fold which is strikingly different from that of animal TGases. However, the secondary structure elements around the catalytic core display a similarity to the same region of animal TGases (MTG) or to the papain-like superfamily of cysteine proteases (GP42) (129,131). GP42 is a multidomain enzyme whose activity is dependent on calcium. It is also synthesized as a zymogen and proteolytic processing is necessary to activate the enzyme. Although GP42 contains a Cys-His-Asp catalytic triad, its arrangement differs from that of the superfamily of cysteine proteases, suggesting that GP42 represents a case of convergent evolution (129). MTG appears to further deviate from the animal paradigm. With 42 kDa, it is the smallest TGase for which a structure is known (against, for example, the 57 and 77 kDa of the zymogenic forms of GP42 and TGase 3, respectively). It is a single domain protein and it contains a catalytic Cys-Asp diad. Also, MTG is a calcium-independent enzyme and no other cofactors are known to

regulate its activity (131,132). However, MTG is synthesized and secreted in zymogenic form and extracellularly processed. Interestingly, the zymogen TGase from *Streptomyces hygroscopicus* is processed into two different active TGases, TGase A and TGase B. The two TGases display alterations in overall charge and hydrophobicity and in substrate affinity. TGase B is generated by the cleavage of the zymogen by extracellular proteases while TGase A derives from the action of TGase B on the zymogen; TGase A, on the other hand cannot process the zymogen (133). Considering the simpler activation process of MTG it is possible that this dual production of TGases from one zymogenic form may be important to regulate the biological processes in which TGase A and B are involved.

Given the considerably simpler structure and activation of MTG, it is perhaps unexpected that no clear relatedness of this enzyme to the NlpC/P60 domain has been recognized (131). The structural and biochemical studies on both GP42 and MTG suggest that TGases can be divided into two main groups, the animal-type enzymes and those, seemingly unrelated, of different evolutionary origin. However, the lack of representation of non-animal-like TGases in the current databases makes it hard to understand the structural basis and requirements for the protein cross-linking reaction, as well as the structural or evolutionary relationships among unrelated TGases.

As stated above, TGases tend to be present in developmental processes in the creation of stable, long lasting structures. This is related to their protein cross-linking ability, where by creating stable and protease resistant covalent bonds between different proteins, TGases contribute to the chemical and physical resistance of tissues and structures (101). Thus, it's not surprising that such an enzyme is involved in spore coat assembly. Although great progress has been made in our general view of spore coat assembly, detailed studies on the assembly of individual coat proteins is limited. Furthermore, the current information on bacterial TGases is lacking and Tgl is the only enzyme of this group for which a physiological role is defined. Hence, by studying Tgl, we wanted to contribute not only to a better understanding of the mechanisms involved in the assembly of the

coat proteins, but also to the knowledge of the structure and function of bacterial TGases. To that end, we focused on three main objectives:

1) to solve the structure of Tgl and define its active site;
2) to understand how Tgl is assembled onto the spore coat and;
3) to elucidate the role of the enzyme in spore coat morphogenesis.

REFERENCES

1. **Kaiser D, Robinson M, Kroos L.** 2010. Myxobacteria, Polarity, and Multicellular Morphogenesis. Cold Spring Harbor Perspectives in Biology **2:**a000380-a000380.

2. **Hincke MT, Nys Y, Gautron J, Mann K, Rodriguez-Navarro AB, McKee MD.** 2012. The eggshell: structure, composition and mineralization. Front Biosci (Landmark Ed) **17:**1266-1280.

3. **Erhardt M, Namba K, Hughes KT.** 2010. Bacterial Nanomachines: The Flagellum and Type III Injectisome. Cold Spring Harbor Perspectives in Biology **2:**a000299-a000299.

4. **Fokine A, Rossmann MG.** 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage **4:**e28281.

5. **Setlow P.** 2006. Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. Journal of applied microbiology **101:**514-525.

6. **Riesenman PJ, Nicholson WL.** 2000. Role of the spore coat layers in *Bacillus subtilis* spore resistance to hydrogen peroxide, artificial UV-C, UV-B, and solar UV radiation. Applied and environmental microbiology **66:**620-626.

7. **Henriques AO, Moran CP, Jr.** 2007. Structure, assembly, and function of the spore surface layers. Annual review of microbiology **61:**555-588.

8. **McKenney PT, Driks A, Eichenberger P.** 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol **11:**33-44.

9. **Setlow P.** 2003. Spore germination. Current opinion in microbiology **6:**550-556.

10. **Senesi S.** 2004. *Bacillus spores* as probiotic products for human use, p. 131-141. *In* Ricca E, Henriques AO, Cutting SM (ed.), Bacterial spore formers: probiotics and emerging applications. Horizon Bioscience.

11. **Knecht LD, Pasini P, Daunert S.** 2011. Bacterial spores as platforms for bioanalytical and biomedical applications. Anal Bioanal Chem **400:**977-989.

12. **Rodrigues MM, de Souza RD, Batista MT, Luiz WB, Cavalcante RCM, Amorim JH, Bizerra RSP, Martins EG, de Souza Ferreira LC.** 2014. *Bacillus subtilis* Spores as Vaccine Adjuvants: Further Insights into the Mechanisms of Action. PLoS One **9:**e87454.

13. **Ricca E, Cutting SM.** 2003. Emerging Applications of Bacterial Spores in Nanobiotechnology. J Nanobiotechnology **1:**6.

14. **Kim H, Hahn M, Grabowski P, McPherson DC, Otte MM, Wang R, Ferguson CC, Eichenberger P, Driks A.** 2006. The *Bacillus subtilis* spore coat protein interaction network. Molecular microbiology **59:**487-502.

15. **Kuwana R, Kasahara Y, Fujibayashi M, Takamatsu H, Ogasawara N, Watabe K.** 2002. Proteomics characterization of novel spore proteins of *Bacillus subtilis*. Microbiology (Reading, England) **148:**3971-3982.

16. **Lai EM, Phadke ND, Kachman MT, Giorno R, Vazquez S, Vazquez JA, Maddock JR, Driks A.** 2003. Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*. Journal of bacteriology **185:**1443-1454.

17. **Henriques AO, Moran CP, Jr.** 2000. Structure and assembly of the bacterial endospore coat. Methods (San Diego, Calif **20:**95-110.

18. **Henriques AO, T. V. Costa, L. O. Martins, and R. Zilhão.** 2004. Functional architecture and assembly of the spore coat, p. 65-85. *In* Ezio Ricca AOH, and Simon M. Cutting (ed.), Bacterial Spores: Probiotics and Emerging Applications. Horizon Scientific Press, London, UK.

19. **Makarova KS, Aravind L, Koonin EV.** 1999. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. Protein Sci **8:**1714-1719.

20. **Setlow P.** 2014. Germination of Spores of *Bacillus* Species: What We Know and Do Not Know. Journal of bacteriology **196:**1297-1305.

21. **Leggett MJ, McDonnell G, Denyer SP, Setlow P, Maillard JY.** 2012. Bacterial spore structures and their protective role in biocide resistance. Journal of applied microbiology **113:**485-498.

22. **Popham DL.** 2002. Specialized peptidoglycan of the bacterial endospore: the inner wall of the lockbox. Cell Mol Life Sci **59:**426-433.

23. **Westphal AJ, Price PB, Leighton TJ, Wheeler KE.** 2003. Kinetics of size changes of individual *Bacillus thuringiensis* spores in response to changes in relative humidity. Proceedings of the National Academy of Sciences of the United States of America **100:**3461-3466.

24. **Catalano FA, Meador-Parton J, Popham DL, Driks A.** 2001. Amino acids in the *Bacillus subtilis* morphogenetic protein SpoIVA with roles in spore coat and cortex formation. Journal of bacteriology **183:**1645-1654.

25. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2010. Localization of proteins to different layers and regions of *Bacillus subtilis* spore coats. Journal of bacteriology **192:**518-524.

26. **McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z, Eichenberger P.** 2010. A distance-weighted interaction map reveals a previously uncharacterized layer of the *Bacillus subtilis* spore coat. Curr Biol **20:**934-938.

27. **Sanchez-Salas JL, Setlow B, Zhang P, Li Yq, Setlow P.** 2011. Maturation of Released Spores Is Necessary for Acquisition of Full Spore Heat Resistance during *Bacillus subtilis* Sporulation. Applied and environmental microbiology **77:**6746-6754.

28. **Moir A, Corfe BM, Behravan J.** 2002. Spore germination. Cell Mol Life Sci **59:**403-409.

29. **Higgins D, Dworkin J.** 2012. Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol Rev **36:**131-148.

30. **Piggot PJ, Hilbert DW.** 2004. Sporulation of *Bacillus subtilis*. Current opinion in microbiology **7:**579-586.

31. **Vasudevan P, Weaver A, Reichert ED, Linnstaedt SD, Popham DL.** 2007. Spore cortex formation in *Bacillus subtilis* is regulated by accumulation of peptidoglycan precursors under the control of sigma K. Molecular microbiology **65:**1582-1594.

32. **Meador-Parton J, Popham DL.** 2000. Structural analysis of *Bacillus subtilis* spore peptidoglycan during sporulation. Journal of bacteriology **182:**4491-4499.

33. **Kroos L.** 2007. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu Rev Genet **41:**13-39.

34. **McPherson DC, Driks A, Popham DL.** 2001. Two class A high-molecular-weight penicillin-binding proteins of *Bacillus subtilis* play redundant roles in sporulation. Journal of bacteriology **183:**6046-6053.

35. **Popham DL, Gilmore ME, Setlow P.** 1999. Roles of low-molecular-weight penicillin-binding proteins in *Bacillus subtilis* spore peptidoglycan synthesis and spore properties. Journal of bacteriology **181:**126-132.

36. **Piggot PJ, Coote JG.** 1976. Genetic aspects of bacterial endospore formation. Bacteriological reviews **40:**908-962.

37. **Roels S, Driks A, Losick R.** 1992. Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. Journal of bacteriology **174:**575-585.

38. **Takamatsu H, Kodama T, Nakayama T, Watabe K.** 1999. Characterization of the *yrbA* gene of *Bacillus subtilis*, involved in resistance and germination of spores. Journal of bacteriology **181:**4986-4994.

39. **Driks A, Roels S, Beall B, Moran CP, Jr., Losick R.** 1994. Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. Genes & development **8:**234-244.

40. **Zheng LB, Donovan WP, Fitz-James PC, Losick R.** 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. Genes & development **2:**1047-1054.

41. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2011. Proteins involved in formation of the outermost layer of *Bacillus subtilis* spores. Journal of bacteriology **193:**4075-4080.

42. **Wang KH, Isidro AL, Domingues L, Eskandarian HA, McKenney PT, Drew K, Grabowski P, Chua MH, Barry SN, Guan M, Bonneau R, Henriques AO, Eichenberger P.** 2009. The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*. Molecular microbiology **74:**634-649.

43. **Levin PA, Fan N, Ricca E, Driks A, Losick R, Cutting S.** 1993. An unusually small gene required for sporulation by *Bacillus subtilis*. Molecular microbiology **9:**761-771.

44. **Beall B, Driks A, Losick R, Moran CP, Jr.** 1993. Cloning and characterization of a gene required for assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **175:**1705-1716.

45. **McKenney PT, Eichenberger P.** 2012. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. Molecular microbiology **83:**245-260.

46. **Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS biology **2:**e328.

47. **Costa T, Serrano M, Steil L, Volker U, Moran CP, Jr., Henriques AO.** 2007. The timing of cotE expression affects *Bacillus subtilis* spore coat morphology but not lysozyme resistance. Journal of bacteriology **189:**2401-2410.

48. **Zheng LB, Losick R.** 1990. Cascade regulation of spore coat gene expression in *Bacillus subtilis*. Journal of molecular biology **212:**645-660.

49. **Zhang J, Ichikawa H, Halberg R, Kroos L, Aronson AI.** 1994. Regulation of the transcription of a cluster of *Bacillus subtilis* spore coat genes. Journal of molecular biology **240:**405-415.

50. **Ramamurthi KS, Losick R.** 2008. ATP-driven self-assembly of a morphogenetic protein in *Bacillus subtilis*. Mol Cell **31:**406-414.

51. **Little S, Driks A.** 2001. Functional analysis of the *Bacillus subtilis* morphogenetic spore coat protein CotE. Molecular microbiology **42:**1107-1120.

52. **Ozin AJ, Samford CS, Henriques AO, Moran CP, Jr.** 2001. SpoVID guides SafA to the spore coat in *Bacillus subtilis*. Journal of bacteriology **183:**3041-3049.

53. **Krajcikova D, Lukacova M, Mullerova D, Cutting SM, Barak I.** 2009. Searching for protein-protein interactions within the *Bacillus subtilis* spore coat. Journal of bacteriology **191:**3212-3219.

54. **Mullerova D, Krajcikova D, Barak I.** 2009. Interactions between *Bacillus subtilis* early spore coat morphogenetic proteins. FEMS microbiology letters **299:**74-85.

55. **Qiao H, Krajcikova D, Xing C, Lu B, Hao J, Ke X, Wang H, Barak I, Tang J.** 2013. Study of the interactions between the key spore coat morphogenetic proteins CotE and SpoVID. Journal of structural biology **181:**128-135.

56. **Rubio A, Pogliano K.** 2004. Septal localization of forespore membrane proteins during engulfment in *Bacillus subtilis*. The EMBO journal **23:**1636-1646.

57. **Blaylock B, Jiang X, Rubio A, Moran CP, Jr., Pogliano K.** 2004. Zipper-like interaction between proteins in adjacent daughter cells mediates protein localization. Genes & development **18:**2916-2928.

58. **Camp AH, Losick R.** 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*. Genes & development **23:**1014-1024.

59. **Meisner J, Wang X, Serrano M, Henriques AO, Moran CP, Jr.** 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. Proceedings of the National Academy of Sciences of the United States of America **105:**15100-15105.

60. **Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Jr., Rudner DZ.** 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. PLoS Genet **5:**e1000566.

61. **Broder DH, Pogliano K.** 2006. Forespore engulfment mediated by a ratchet-like mechanism. Cell **126:**917-928.

62. **Ramamurthi KS, Clapham KR, Losick R.** 2006. Peptide anchoring spore coat assembly to the outer forespore membrane in *Bacillus subtilis*. Molecular microbiology **62:**1547-1557.

63. **Ramamurthi KS, Lecuyer S, Stone HA, Losick R.** 2009. Geometric cue for protein localization in a bacterium. Science (New York, N.Y **323:**1354-1357.

64. **de Francesco M, Jacobs JZ, Nunes F, Serrano M, McKenney PT, Chua MH, Henriques AO, Eichenberger P.** 2012. Physical interaction between coat morphogenetic proteins SpoVID and CotE is necessary for spore encasement in *Bacillus subtilis*. Journal of bacteriology **194:**4941-4950.

65. **Bauer T, Little S, Stover AG, Driks A.** 1999. Functional regions of the *Bacillus subtilis* spore coat morphogenetic protein CotE. Journal of bacteriology **181:**7043-7051.

66. **Costa T, Isidro AL, Moran CP, Jr., Henriques AO.** 2006. Interaction between coat morphogenetic proteins SafA and SpoVID. Journal of bacteriology **188:**7731-7741.

67. **Ozin AJ, Costa T, Henriques AO, Moran CP, Jr.** 2001. Alternative translation initiation produces a short form of a spore coat protein in *Bacillus subtilis*. Journal of bacteriology **183:**2032-2040.

68. **Ozin AJ, Henriques AO, Yi H, Moran CP, Jr.** 2000. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **182:**1828-1833.

69. **Qiao H, Krajcikova D, Liu C, Li Y, Wang H, Barak I, Tang J.** 2012. The interactions of spore-coat morphogenetic proteins studied by single-molecule recognition force spectroscopy. Chem Asian J **7:**725-731.

70. **Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT.** 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic acids research **41:**W349-357.

71. **Li X, Romero P, Rani M, Dunker AK, Obradovic Z.** 1999. Predicting Protein Disorder for N-, C-, and Internal Regions. Genome Inform Ser Workshop Genome Inform **10:**30-40.

72. **Romero, Obradovic, Dunker K.** 1997. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome Inform Ser Workshop Genome Inform **8:**110-124.

73. **Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK.** 2001. Sequence complexity of disordered protein. Proteins **42:**38-48.

74. **Coote JG.** 1972. Sporulation in *Bacillus subtilis*. Characterization of oligosporogenous mutants and comparison of their phenotypes with those of asporogenous mutants. J Gen Microbiol **71:**1-15.

75. **Ebmeier SE, Tan IS, Clapham KR, Ramamurthi KS.** 2012. Small proteins link coat and cortex assembly during sporulation in *Bacillus subtilis*. Molecular microbiology **84:**682-696.

76. **Takamatsu H, Chikahiro Y, Kodama T, Koide H, Kozuka S, Tochikubo K, Watabe K.** 1998. A spore coat protein, CotS, of *Bacillus subtilis* is synthesized under the regulation of sigmaK and GerE during development and is located in the inner coat layer of spores. Journal of bacteriology **180:**2968-2974.

77. **Costa T, Steil L, Martins LO, Volker U, Henriques AO.** 2004. Assembly of an oxalate decarboxylase produced under sigmaK control into the *Bacillus subtilis* spore coat. Journal of bacteriology **186:**1462-1474.

78. **Zilhao R, Naclerio G, Henriques AO, Baccigalupi L, Moran CP, Jr., Ricca E.** 1999. Assembly requirements and role of CotH during spore coat formation in *Bacillus subtilis*. Journal of bacteriology **181:**2631-2633.

79. **McPherson DC, Kim H, Hahn M, Wang R, Grabowski P, Eichenberger P, Driks A.** 2005. Characterization of the *Bacillus subtilis* spore morphogenetic coat protein CotO. Journal of bacteriology **187:**8278-8290.

80. **Naclerio G, Baccigalupi L, Zilhao R, De Felice M, Ricca E.** 1996. *Bacillus subtilis* spore coat assembly requires cotH gene expression. Journal of bacteriology **178:**4375-4380.

81. **Driks A, Isticato R, Sirec T, Giglio R, Baccigalupi L, Rusciano G, Pesce G, Zito G, Sasso A, De Felice M, Ricca E.** 2013. Flexibility of the Programme of Spore Coat Formation in *Bacillus subtilis*: Bypass of CotE Requirement by Over-Production of CotH. PLoS One **8:**e74949.

82. **Zilhao R, Serrano M, Isticato R, Ricca E, Moran CP, Jr., Henriques AO.** 2004. Interactions among CotB, CotG, and CotH during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **186:**1110-1119.

83. **Takamatsu H, Kodama T, Watabe K.** 1999. Assembly of the CotSA coat protein into spores requires CotS in *Bacillus subtilis*. FEMS microbiology letters **174:**201-206.

84. **Seyler RW, Jr., Henriques AO, Ozin AJ, Moran CP, Jr.** 1997. Assembly and interactions of *cotJ*-encoded proteins, constituents of the inner layers of the *Bacillus subtilis* spore coat. Molecular microbiology **25:**955-966.

85. **Ragkousi K, Eichenberger P, van Ooij C, Setlow P.** 2003. Identification of a new gene essential for germination of *Bacillus subtilis* spores with Ca2+-dipicolinate. Journal of bacteriology **185:**2315-2329.

86. **Ishikawa S, Yamane K, Sekiguchi J.** 1998. Regulation and characterization of a newly deduced cell wall hydrolase gene (*cwlJ*) which affects germination of *Bacillus subtilis* spores. Journal of bacteriology **180:**1375-1380.

87. **McPherson SA, Li M, Kearney JF, Turnbough CL, Jr.** 2010. ExsB, an unusually highly phosphorylated protein required for the stable attachment of the exosporium of *Bacillus anthracis*. Molecular microbiology **76:**1527-1538.

88. **Takamatsu H, Kodama T, Imamura A, Asai K, Kobayashi K, Nakayama T, Ogasawara N, Watabe K.** 2000. The *Bacillus subtilis yabG* gene is transcribed by SigK RNA polymerase during sporulation, and *yabG* mutant spores have altered coat protein composition. Journal of bacteriology **182:**1883-1888.

89. **Takamatsu H, Imamura A, Kodama T, Asai K, Ogasawara N, Watabe K.** 2000. The *yabG* gene of *Bacillus subtilis* encodes a sporulation specific protease which is involved in the processing of several spore coat proteins. FEMS microbiology letters **192:**33-38.

90. **Pandey NK, Aronson AI.** 1979. Properties of the *Bacillus subtilis* spore coat. Journal of bacteriology **137:**1208-1218.

91. **Driks A.** 1999. *Bacillus subtilis* spore coat. Microbiol Mol Biol Rev **63:**1-20.

92. **Katsunori Kobayashi YK, Kiyoshi Miwa, Shigeru Yamanaka.** 1996. ε-(γ-Glutamyl)lysine cross-links of spore coat proteins and transglutaminase activity in *Bacillus subtilis*. FEMS microbiology letters **144:**157-160.

93. **Ragkousi K, Setlow P.** 2004. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. Journal of bacteriology **186:**5567-5575.

94. **Suzuki S, Izawa Y, Kobayashi K, Eto Y, Yamanaka S, Kubota K, Yokozeki K.** 2000. Purification and characterization of novel transglutaminase from *Bacillus subtilis* spores. Bioscience, biotechnology, and biochemistry **64:**2344-2351.

95. **Kuwana R, Okuda N, Takamatsu H, Watabe K.** 2006. Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. Journal of biochemistry **139:**887-901.

96. **Steil L, Serrano M, Henriques AO, Volker U.** 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. Microbiology (Reading, England) **151:**399-420.

97. **Moir A.** 1981. Germination properties of a spore coat-defective mutant of *Bacillus subtilis*. Journal of bacteriology **146:**1106-1116.

98. **Zilhao R, Isticato R, Martins LO, Steil L, Volker U, Ricca E, Moran CP, Jr., Henriques AO.** 2005. Assembly and function of a spore coat-associated transglutaminase of *Bacillus subtilis*. Journal of bacteriology **187:**7753-7764.

99. **Ghosh S, Setlow B, Wahome PG, Cowan AE, Plomp M, Malkin AJ, Setlow P.** 2008. Characterization of spores of *Bacillus subtilis* that lack most coat layers. Journal of bacteriology **190:**6741-6748.

100. **Monroe A, Setlow P.** 2006. Localization of the transglutaminase cross-linking sites in the *Bacillus subtilis* spore coat protein GerQ. Journal of bacteriology **188:**7609-7616.

101. **Lorand L, Graham RM.** 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature reviews **4:**140-156.

102. **Placido D, Fernandes CG, Isidro A, Carrondo MA, Henriques AO, Archer M.** 2008. Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. Protein expression and purification **59:**1-8.

103. **Kobayashi K, Suzuki SI, Izawa Y, Miwa K, Yamanaka S.** 1998. Transglutaminase in sporulating cells of *Bacillus subtilis*. J Gen Appl Microbiol **44:**85-91.

104. **Griffin M, Casadio R, Bergamini CM.** 2002. Transglutaminases: nature's biological glues. The Biochemical journal **368:**377-396.

105. **Beninati S, Piacentini M.** 2004. The transglutaminase family: an overview: minireview article. Amino acids **26:**367-372.

106. **Ahvazi B, Boeshans KM, Rastinejad F.** 2004. The emerging structural understanding of transglutaminase 3. Journal of structural biology **147:**200-207.

107. **Boeshans KM, Mueser TC, Ahvazi B.** 2007. A three-dimensional model of the human transglutaminase 1: insights into the understanding of lamellar ichthyosis. Journal of molecular modeling **13:**233-246.

108. **Nemes Z, Marekov LN, Fesus L, Steinert PM.** 1999. A novel function for transglutaminase 1: attachment of long-chain omega-hydroxyceramides to involucrin by ester bond formation. Proceedings of the National Academy of Sciences of the United States of America **96:**8402-8407.

109. **Gundemir S, Colak G, Tucholski J, Johnson GVW.** 2012. Transglutaminase 2: A molecular Swiss army knife. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1823:**406-419.

110. **Bozzini S, Giuliano L, Altomare L, Petrini P, Bandiera A, Conconi M, Farè S, Tanzi M.** 2011. Enzymatic cross-linking of human recombinant elastin (HELP) as biomimetic approach in vascular tissue engineering. J Mater Sci: Mater Med **22:**2641-2650.

111. **Mero A, Spolaore B, Veronese FM, Fontana A.** 2009. Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. Bioconjugate chemistry **20:**384-389.

112. **Lin CW, Ting AY.** 2006. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins *in vitro* and on the surface of living cells. Journal of the American Chemical Society **128:**4542-4543.

113. **Kim SY, Jeitner TM, Steinert PM.** 2002. Transglutaminases in disease. Neurochemistry international **40:**85-103.

114. **Klöck C, Khosla C.** 2012. Regulation of the activities of the mammalian transglutaminase family of enzymes. Protein Science **21:**1781-1791.

115. **Greenberg CS, Birckbichler PJ, Rice RH.** 1991. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. Faseb J **5:**3071-3077.

116. **Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE.** 1994. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Sci **3:**1131-1135.

117. **Anantharaman V, Aravind L.** 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome biology **4:**R11.

118. **Xu Q, Sudek S, McMullan D, Miller MD, Geierstanger B, Jones DH, Krishna SS, Spraggon G, Bursalay B, Abdubek P, Acosta C, Ambing E, Astakhova T, Axelrod HL, Carlton D, Caruthers J, Chiu HJ, Clayton T, Deller MC, Duan L, Elias Y, Elsliger MA, Feuerhelm J, Grzechnik SK, Hale J, Won Han G, Haugen J, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Kumar A, Marciano D, Morse AT, Nigoghossian E, Okach L, Oommachen S, Paulsen J, Reyes R, Rife CL, Trout CV, van den Bedem H, Weekes D, White A, Wolf G, Zubieta C, Hodgson KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA.** 2009. Structural basis of murein peptide specificity of a gamma-D-glutamyl-l-diamino acid endopeptidase. Structure **17:**303-313.

119. **Aramini JM, Rossi P, Huang YJ, Zhao L, Jiang M, Maglaqui M, Xiao R, Locke J, Nair R, Rost B, Acton TB, Inouye M, Montelione GT.** 2008. Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from *Escherichia coli*: structural evidence for a novel cysteine peptidase catalytic triad. Biochemistry **47:**9715-9717.

120. **Pinkas DM, Strop P, Brunger AT, Khosla C.** 2007. Transglutaminase 2 undergoes a large conformational change upon activation. PLoS biology **5:**e327.

121. **Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM.** 2002. Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. The EMBO journal **21:**2055-2067.

122. **Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM.** 2003. Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. The Journal of biological chemistry **278:**23834-23841.

123. **Stieler M, Weber J, Hils M, Kolb P, Heine A, Büchold C, Pasternack R, Klebe G.** 2013. Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. Angewandte Chemie International Edition **52:**11930-11934.

124. **Park D, Choi SS, Ha K-S.** 2010. Transglutaminase 2: a multi-functional protein in multiple subcellular compartments. Amino acids **39:**619-631.

125. **Lorand L.** 2001. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. Ann N Y Acad Sci **936:**291-311.

126. **Yang Z, Mochalkin I, Doolittle RF.** 2000. A model of fibrin formation based on crystal structures of fibrinogen and fibrin fragments complexed with synthetic peptides. Proceedings of the National Academy of Sciences of the United States of America **97:**14156-14161.

127. **Steinert PM, Chung SI, Kim SY.** 1996. Inactive zymogen and highly active proteolytically processed membrane-bound forms of the transglutaminase 1 enzyme in human epidermal keratinocytes. Biochemical and biophysical research communications **221:**101-106.

128. **Eckert RL, Sturniolo MT, Jans R, Kraft CA, Jiang H, Rorke EA.** 2008. TIG3: a regulator of type I transglutaminase activity in epidermis. Amino acids **36:**739-746.

129. **Reiss K, Kirchner E, Gijzen M, Zocher G, Loffelhardt B, Nurnberger T, Stehle T, Brunner F.** 2011. Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria. Journal of Biological Chemistry **286:**42585-42593.

130. **Yang MT, Chang CH, Wang JM, Wu TK, Wang YK, Chang CY, Li TT.** 2010. Crystal Structure and Inhibition Studies of Transglutaminase from *Streptomyces mobaraense*. Journal of Biological Chemistry **286:**7301-7307.

131. **Kashiwagi T, Yokoyama K, Ishikawa K, Ono K, Ejima D, Matsui H, Suzuki E.** 2002. Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. The Journal of biological chemistry **277:**44252-44260.

132. **Pasternack R, Dorsch S, Otterbach JT, Robenek IR, Wolf S, Fuchsbauer HL.** 1998. Bacterial pro-transglutaminase from *Streptoverticillium mobaraense*-purification, characterisation and sequence of the zymogen. European journal of biochemistry / FEBS **257:**570-576.

133. **Cui W, Yang X, Fang Y, Zhou S, Liu S, Du G, Du K, Chen J, Tao G, Zhou Z.** 2013. Discovery of two transglutaminases derived from same zymogen for the *Streptomyces hygroscopicus* and analysis of their formation processes. Journal of the Science of Food and Agriculture **93:**1711-1717.

Chapter 2

Structural and Functional Characterization of a Streamlined Bacterial Transglutaminase

This Chapter contains data to be published in: Catarina G. Fernandes, Diana Plácido, Diana Lousa, José A. Brito, Anabela Isidro, Cláudio M. Soares, Jan Pohl, Maria A. Carrondo, Margarida Archer and Adriano O. Henriques. 2014. *Structural and Functional Characterization of a Streamlined Bacterial Transglutaminase*. To be submitted to The Journal of Biological Chemistry.

The Tgl structure was solved by Diana Plácido and José A. Brito, from the Membrane Protein Crystallography Group, ITQB. The molecular dynamics simulations were conducted by Diana Lousã and Claúdio Soares of the Protein Modeling Group, ITQB. Mass spectrometry and N-terminal analysis were performed by Jan Pohl, Center for Disease Control and Prevention, Atlanta, USA. The remaining work was performed by the author of this dissertation.

ABSTRACT

Transglutaminases (TGases) catalyze protein cross-linking reactions imparting chemical and physical resilience to cellular structures or tissues. We report the crystal structure and functional analysis of a bacterial TGase (Tgl) in native form, and in a complex with the inhibitor cysteamine. Tgl cross-links the bacterial spore surface and is the smallest TGase characterized to date. The catalytic core of Tgl is related to the cysteine endopeptidase NlpC/P60 domain, thought to be close to the ancestral unit of the papain-like fold, which includes TGases. We found that Tgl functions through a unique dual catalytic diad formed by Cys116 and an acidic residue, Glu187, which can be non-reciprocally substituted by Glu115. Strikingly, the active site Cys is insulated within a hydrophobic tunnel that traverses the molecule from side to side. Evidence suggests that residues at both entrances of the tunnel are involved in substrate recognition and the tunnel itself, is most likely, a critical determinant for enzymatic specificity. No factors are known that control the activity of Tgl, and protein labeling and cross-linking was possible under simple reaction conditions. We propose that Tgl approaches the necessary and sufficient structural characteristics for the protein cross-linking reaction catalyzed by TGases.

INTRODUCTION

Transglutaminases (TGases) are best known for their transamidation activity, specifically their ability to catalyze formation of cross-links between peptide-bound glutaminyl and lysyl residues (Q and K substrates, respectively) (1-3). The resulting ε-(γ-glutamyl)lysyl isopeptide cross-link is stable and protease resistant, and contributes to the chemical and mechanical resistance of several protein assemblies, cellular structures, or tissues (3). Although not all the biological functions of TGases require catalytic activity, protein cross-linking is central to a variety of important processes including clotting phenomena, assembly of the extracellular matrix, hair and skin formation, or hardening of the fertilization membrane (1-4). In another type of transamidation reaction, TGases react with primary amines present in simple molecules or naturally occurring polyamines, leading to amine

incorporation (1-3). TGase-catalyzed hydrolytic reactions have also been reported (5). In recent years, TGases have been exploited in applications ranging from the construction of matrices for regenerative medicine (6) to the site-specific modification of proteins (7,8).

Structural and biochemical studies have shown that animal TGases and the papain-like cysteine proteases share a similar active site arrangement and catalytic acylation-deacylation pathway, and are grouped under the same superfamily (9). All animal TGases that have been characterized, of which the human enzymes have been the most extensively studied, employ a charge-relay Cys-His-Asp catalytic triad, and the reaction mechanism can be described as two successive nucleophilic attacks (Figure 2.1). First, the thiolate anion of the catalytic Cys attacks the amide group at the γ position of a Gln side chain, leading to the formation of a covalent acyl-enzyme intermediate (Figure 2.1a-c). In a second step, the acyl-enzyme intermediate is attacked by a primary amine generating the final product (Figure 2.1d-f). During the reaction, a thiolate/imidazolium ion pair is formed between the catalytic Cys and His residues, the latter acting both as proton donor and acceptor during catalysis (Figure 2.1b,d). An hydrogen bond between the Asp and His residues of the catalytic triad, facilitates the correct orientation of the basic amino acid (Figure 2.1) (9-11). Contrary to the papain-like proteases, in which the second substrate is an activated water molecule, exclusion of water from the active site has been postulated to be a crucial requirement for TGases, so that transamidation is favored over deamination reactions, in which Gln residues are converted to Glu (12). Adding to the mechanistic similarity, the catalytic core of TGases appears to derive from the minimal ancestral structural unit of the thiol-protease fold, whose closest current representative may correspond to the NlpC/P60 domain (13). This fold derives its name from the NlpC/P60-type cell wall endopeptidases, which also show a catalytic triad composed of Cys, His, and a third polar residue (13,14).

Animal TGases are large, complex multidomain enzymes, whose activity is tightly regulated. They are activated by calcium and either produced in zymogenic form, bound by inhibitory subunits, and/or negatively modulated by GTP/GDP or ATP (3,15,16). The complexity of the animal enzymes most

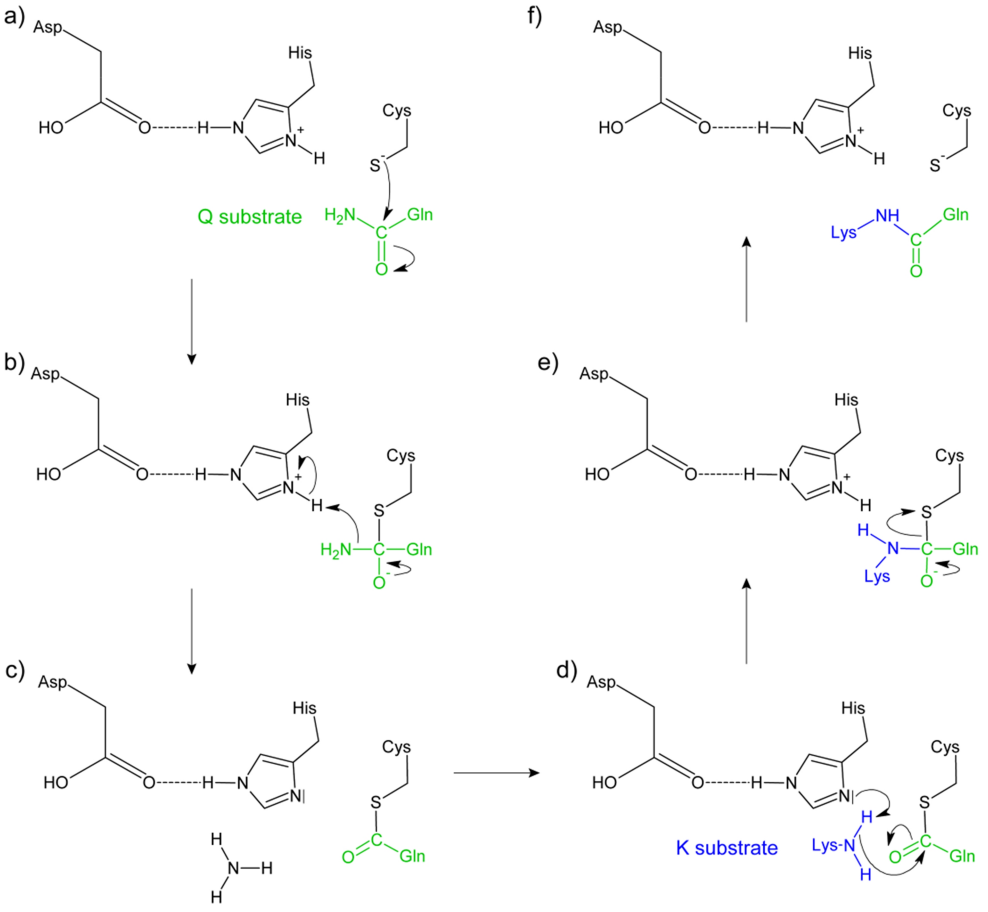


Figure 2.1. **Proposed reaction mechanism for the animal TGases.** The reaction mechanism was proposed based on the similarities between the catalytic cores of TGases and cysteine proteases (9). Gln and Lys refer to the protein-bound Gln and Lys residues. The reaction mechanism involves 6 different steps: a), the thiolate ion nucleophilically attacks the acyl donor, the side chain of the Gln residue, which results in an oxyanion intermediate, as depicted in b); the regeneration of the carbonyl group of the tetrahedral intermediate leads to the release of an ammonia molecule (in c); d), the catalytic His residue, removes a proton from the primary amine of the Lys residue; this leads to the nucleophilic attack of the Lys residue to the carbonyl group of the intermediate product formed in c, and to the formation of a second tetrahedral intermediate (showed in e); f), as the oxyanion intermediate regenerates the carbonyl group, the thioester bond is cleaved re-generating the initial catalytic core and releasing the cross-linked protein product. In MTG, which displays an inversion of the His and Asp residues relative to the animal-type TGases, the role of His is played by an Asp, while the His residue is not essential for catalysis (26).

likely reflects their multiple functions and interactions *in vivo* (15,17,18). For example, human tissue TGase, or TGase 2, a 80 kDa enzyme found in many cell types, has enzymatic as well as transamidase-independent functions, both inside and outside of the cell (15,19). Like the other animal TGases, TGase 2 is formed by four domains, which in the absence of Ca^{2+} assume a

compact, transamidase-inactive conformation (20) (Figure 1.5). Activation of TGase 2, as revealed by its structure in a complex with an inhibitor that mimics a natural substrate, involves a large conformational rearrangement, upon which the enzyme acquires an extended ellipsoidal structure (20). This rearrangement exposes a hydrophobic tunnel leading to the active site (20). A tunnel is also seen in the active forms of human TGase 3 and Factor XIIIa (21,22). Importantly, the active site is inaccessible in the inactive forms of the enzymes (9,20,21,23). Similar alterations may be involved in the activation of other animal-like enzymes (10).

Database searches have led to the identification of proteins homologous to animal TGases in all archaea, in some bacteria and, in eukaryotes (in yeast species and the nematode *Caenorhabditis elegans*) implying that animal-like TGases are widespread (24). The sequence similarity of the different homologs of this superfamily is centered around the three predicted catalytic residues (24). Perhaps reflecting different substrate specificities and biological processes in which these putative TGases intervene, some have signal peptides or transmembrane domains, while others appear cytoplasmic, and for some of the predicted intracellular enzymes, additional protein-protein interaction domains, such as SH3 or C4-type Zn-fingers, are present (24). Interestingly, some of the putative TGases had already been shown to be proteases, in agreement with the notion that animal-like TGases have evolved from an ancestral protease (24).

While animal-like TGases can be detected through sequence-based searches, some already characterized TGases show no sequence similarity to animal TGases (25,26) which suggests the existence of a distinct group of non-animal-like TGases. As such it is hard to assess to what extent the mechanistic and structural features of animal TGases are conserved among this group of enzymes.

So far, the only structure available of a bacterial TGase is that of MTG, isolated from the bacterium *Streptomyces mobaraensis* (26,27). MTG is also produced as a zymogen, but the enzyme shows several deviations from the animal TGase paradigm: it lacks similarity, at the primary and tertiary structure level, when compared to animal TGases, it has a much smaller size (38

kDa), it's Ca^{2+}-independent and employs a Cys-Asp catalytic diad (26).

Here, we have studied a second bacterial TGase, called Tgl, and we present the crystal structure and functional analysis of this enzyme. Tgl shows no sequence conservation with MTG or other TGases, and is unique to a deeply-rooted group of spore-forming microorganisms, which includes *Bacillus* and *Clostridium* species, suggesting that it is an ancient enzyme (28). Tgl is produced during endospore differentiation and participates in the cross-linking of a multiprotein surface structure known as the coat that encases and protects the spore (29). Tgl is produced in active form, and no factor (Ca^{2+}, GTP or other) is known to control its activity (30). At 28 kDa, Tgl is also the smallest TGase that has been characterized to date. We show here that Tgl has a unique, partially redundant Cys116-Glu187/Glu115 catalytic diad. We also show that Tgl has an NlpC/P60 domain at its catalytic core and that the catalytic Cys116 residue is located within a tunnel, insulated from water, which may confer strict specificity to the enzyme. Tgl appears to be a streamlined TGase and its characterization offers insight into the minimal structural requirements for protein cross-linking.

MATERIALS AND METHODS

***Point mutations in* tgl.** Mutations leading to single alanine substitutions were introduced in the *tgl* gene by PCR using pLOM4 as the template (31). pLOM4 is a pET30a(+) (Novagen) derivative where the expression of tgl_{wt}-his_6 is under the control of the T7*lac* promoter. The C116A, H200A and E115A codon substitutions were introduced in *tgl* using, respectively, primer pairs *tgl*-116D/*tgl*-116R, *tgl*H200Adir/*tgl*H200Arev and Tgl E115Adir/Tgl E115Arev which yielded plasmids pCF1, pCF2 and pCF43 (the sequence of the primers used in this study is given in Table 2.1). Additionally, using pRZ80 (31) as the template and primers *tgl*E187Adir and *tgl*E187Arev, the Glu codon 187 of *tgl* was changed to Ala. The resulting plasmid, pCF5, was digested with *Bgl*II and *EcoR*I to release a 579 bp DNA fragment, which was inserted between the same sites of pLOM4, creating pCF12. The presence of all desired mutations as well as the absence of unwanted mutations was confirmed by DNA sequencing.

Table 2.1. Oligonucleotide primers used in this study.

Name	sequence (5' to 3')
tgl-116D	CCCGTTTTATGCGTTTGAAGCGGCAACCGCAATCGTTATC
tgl-116R	GATAACGATTGCGGTTGCCGCTTCAAACGCATAAAACGGG
Tgl E115Adir	CGGCCCGTTTTATGCGTTTGCATGCGCAACCGCAATCG
Tgl E115Arev	CGATTGCGGTTGCGCATGCAAACGCATAAAACGGGCCG
*tgl*E187Adir	CCGCAAAAGGCGCAATGGAGAGGCGCAAATGTGATACTACTGGGGGAAG
*tgl*E187Arev	CTTCCCCCAGTAGTATCACATTTGCGCCTCTCCATTGCGCCTTTTGCGG
*tgl*H200Adir	GGGGAAGATAAATATTTTGCCGCTGGTCTTGGAATCTTAAACGG
*tgl*H200Arev	CCGTTTAAGATTCCAAGACCAGCGGCAAAATATTTATCTTCCCC

The nucleotides altered to introduce single alanine substitutions in Tgl are underlined.

Table 2.2. Bacterial strains and plasmids used in this study.

Strain	Description	Origin
AH4567	*E. coli* DH5α with pCF5, tgl_{E187A}, Amp^R	This study
AH4611	*E. coli* BL21(DE3) with pLOM4 (31), tgl_{wt}-his_6, Kan^R	This study
AH4612	*E. coli* BL21(DE3) with pCF1, tgl_{C116A}-his_6, Kan^R	This study
AH4614	*E. coli* BL21(DE3) with pCF12, tgl_{E187A}-his_6, Kan^R	This study
AH4616	*E. coli* BL21(DE3) with pCF2, tgl_{H200A}-his_6, Kan^R	This study
AH10140	*E. coli* BL21(DE3) with pTC49 (32), *gst-spoVID*$_{202}$, Kan^R	This study
AH10211	*E. coli* BL21(DE3) with pCF43, tgl_{E115A}-his_6, Amp^R	This study

Amp^R, resistance to ampicillin; Kan^R, resistance to kanamycin.

Production and purification of Tgl. The pET30a(+) derivatives (above) were used to overproduce the wild type and the Tgl^{E115A}, Tgl^{C116A}, Tgl^{E187A} and Tgl^{H200A} forms of Tgl as C-terminal fusions to the His_6 tag. All plasmids were introduced into *Escherichia coli* BL21(DE3), creating strains AH4611, AH4612, AH4614, AH4616 and AH10211, respectively (Table 2.2). The wild type and mutant forms of Tgl-His_6 were overproduced using a modified auto-induction protocol (33). Briefly, 100 ml of LB medium supplemented with 100 µg/ml of kanamycin, 1 mM $MgSO_4$, 2 ml of 50x5052 solution (25% glycerol, 2.5% glucose and 10% α-lactose [w/v]) and 5 ml of 20xNPS solution (0.5 M $[NH_4]_2SO_4$, 1 M KH_2PO_4, and 1 M Na_2HPO_4, pH 6.75) were inoculated and incubated at 37° C for about 18 hr with orbital shaking. Tgl was then purified at pH 10 on His-Trap Ni^{2+}-NTA columns (Qiagen), as previously

described (30), and collected in 500 µl fractions. For the activity assays (see below), the fractions with the purified protein were dialyzed overnight against 0.1 M Tris-HCl, pH 8.0, using a 10 kDa cut-off membrane (SnakeSkin, from Pierce). The final concentration of the wild type or mutant forms of Tgl after dialysis was between 4 and 8 mg/ml.

Crystallization, X-ray data collection, phasing and refinement. Tgl was crystallized as previously described (30). Briefly, initial crystallization trials were set up in 96-well sitting-drop plates (Greiner), using a Mini-bee Cartesian robot at 20° C and several commercially available screens. Optimized crystallization conditions consisted of 41% (w/v) PEG 400, 0.2 M Li_2SO_4, and 0.1 M sodium citrate pH 4.0. Noteworthy, no Tgl crystals were formed at pH higher than 4.5-5. Transfer of crystals to solutions of higher pH led to their degradation, even with crystals cross-linked with glutaraldehyde. Tgl:Gd derivative crystals were obtained by soaking, for 30 min, freshly prepared native crystals in the crystallization condition supplemented with 0.1 M Gd-DTPA-BMA. Tgl:cysteamine crystals were prepared by overnight soaking in a 10 mM cystamine solution. Crystals were directly pooled from the crystallization droplets and flash-cooled in liquid nitrogen. X-ray diffraction data was measured at cryogenic temperatures at synchrotron sources, processed with XDS and converted to mtz format with XDSCONV (34). The Tgl structure was solved by single isomorphous replacement with anomalous scattering (SIRAS) method using the Tgl:Gd-derivative (Table 2.3). The autoSHARP module was used within SHARP package (35) and SHELXD (36) located an initial set of 5 heavy atom sites from the putative derivative showing correlation coefficients of 38.1% (CC_{all}), and 18.7% (CC_{weak}), and a PATFOM value of 5.75. After a few rounds of residual map interpretation with SHARP (35) and addition of new derivative sites, an initial electron density map was calculated with the available phases. Final phasing statistics yielded an anomalous phasing power of 0.834 (dropped below 1 at 4.27 Å), and mean figures-of-merit of 19.0% and 17% for acentric and centric reflections, respectively. Phases were improved by density modification within SOLOMON (37) and automated model building was performed with BUCCANEER (38). An initial model, comprising 501 residues in 8 chains which belong to two molecules

in the asymmetric unit, was built and improved with iterative cycles of manual model building with COOT (39) and refinement. Initial refinement rounds were performed with BUSTER-TNT (40) using the macro that accounts for missing parts of the model and final rounds were performed with Phenix (41). The structures of apo-Tgl and Tgl:cysteamine complex are very similar (RMSD for all atoms is 0.29 Å). The stereochemistry of the refined model was assessed with Procheck/ MolProbity (42,43) (Table 2.3). Structure figures were rendered with Pymol (44).

Production and purification of GST-SpoVID$_{202}$. GST-SpoVID$_{202}$ (32) was overproduced in *E. coli* BL21(DE3) (strain AH10140) by auto-induction, as described above for Tgl (except that kanamycin was substituted by 100 µg/ml of ampicillin). To purify GST-SpoVID$_{202}$ the cells were collected by centrifugation (15300 x g for 5 min at 4° C) and resuspended, in 1/10 of the initial culture volume, with VPEX-100 buffer (100 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.1% Triton X-100, 10% glycerol, pH 8.0) supplemented with protease inhibitors (Roche) and benzonase (Novagen). The cells were lysed in a French Pressure cell (at 18000 lb/in^2), the cell debris removed by centrifugation (39200 x g for 1 hr at 4° C), and the supernatant applied to a 1 ml gluthathione sepharose 4B column (GE Healthcare) equilibrated with VPEX-100. The column was washed with 10 volumes of VPEX-200 (same as VPEX-100 but with 200 mM NaCl) followed by 10 volumes of 50 mM Tris-HCl, pH 8.0. Finally, GST-SpoVID$_{202}$ was eluted with 50 mM Tris-HCl, 10 mM gluthathione, pH 8.0, and collected in 500 µl fractions. The purified fractions were dialyzed as described above for Tgl.

Dansylcadaverine labeling assays. Amine incorporation activity was evaluated by monitoring labeling of BSA (New England Biolabs) by dansylcadaverine (dansyl-cd, Fluka) (45). Reaction mixtures containing Tgl (16 µM), BSA (60 µM) and dansyl-cd (0.5 mM) in Tris-HCl 0.1 M, pH 8.0 were incubated at 50° C, samples withdrawn along time and resolved in a 10% SDS-PAGE gel. Labeling of BSA by dansyl-cd was recorded in a UV transilluminator (Chemidoc XRS, Biorad) and quantified using *ImageJ 1.37v* (46). A sample collected following 120 min of reaction with wild type Tgl was loaded onto each gel for the normalization of the fluorescence signal among different assays. All Tgl forms were purified and assayed at least

three times independently.

Cross-linking assays. The wild type or mutant forms of Tgl (16 µM) were incubated with GST-SpoVID$_{202}$ (60 µM) at 50° C in Tris-HCl 0.1 M, pH 8.0, and samples collected at different times for analysis on 10% SDS-PAGE gels. As controls, GST-SpoVID$_{202}$ and Tgl were incubated alone, and samples taken at the beginning and at the end of the assays. All forms of Tgl were assayed independently a minimum of three times.

Mass spectrometry. The products of Tgl cross-linking were resolved by SDS-PAGE and subjected to trypsin digestion and analyzed by nanoLC-ESI-MS/MS on a maXis mass spectrometer (Bruker Daltonics). Briefly, following in-gel digestion of the high molecular weight species resulting from incubation of GST-SpoVID$_{202}$ with Tgl^{wt}, Tgl^{E187A} and Tgl^{H200A}, the extracted peptides were fractionated by nanobore U-HPLC (nanoRSLC, Dionex) on 75 um x 250 mm PepMap120 Å C18 columns (Dionex) eluted at 200 ml/min with acetonitrile gradient in 0.1% aqueous trifluoroacetic acid and analyzed on-line in a Bruker Daltonics model maXis Q-TOF tandem mass spectrometer equipped with a nanoelectrospray source (47). Positive protein identifications of GST-SpoVID$_{202}$ and Tgl were performed using the Mascot algorithm (Matrix Science). Alternatively, the Coomassie-stained protein bands corresponding to GST-SpoVID$_{202}$, Tgl^{wt}, and the high molecular weight species were obtained by SDS-PAGE and electroblotting onto PVDF membranes. The membranes were subjected to N-terminal sequence analysis using model cLC-Procise sequencer (Applied Biosystems). Sequencing afforded unambiguous identification of the first 10 residues. In the case of GST-SpoVID$_{202}$ and Tgl^{wt}, their expected N-terminal sequences were confirmed. In the case of high molecular weight species, N-terminal sequences of both GST-SpoVID$_{202}$ and Tgl^{wt} were identified in each band.

pKa predictions. The predictions of the p*K*a values of all the ionizable residues of Tgl employed a methodology based on continuum electrostatic (running MEAD version 2.2.5) (48,49) and Monte Carlo simulations (using the program PETIT) (50), considering a dielectric constant of 15 for the protein interior (50). This method was benchmarked against a large set of proteins (51) and showed that the overall prediction had an RMSD of 0.84

pK units. The X-ray structure of Tgl co-crystallized with cysteamine was used in the calculations, considering the two monomers found in the crystal (A and B). The p*K*a values obtained correspond to the average over the values obtained for each structure.

Molecular dynamics simulations. Molecular dynamics (MD) simulations were performed with the GROMACS package (52), version 4.0.4 (53), using the GROMOS 53A6 force field (54), and applying a 2 fs integration time step. The SPC model was used for water molecules and the system was simulated under periodic boundary conditions. The Berendsen temperature and pressure coupling algorithms were used to maintain the temperature and pressure constant at 300 K and 1 atm, respectively. A cutoff of 14 Å was applied for Van der Waals and electrostatic interactions and the treatment of long-range electrostatic interactions was done using the particle mesh Ewald method (55) with a cutoff of 9 Å. The neighbor list was updated every 10 steps. The protein bond lengths were constrained with the LINCS (56) algorithm and SETTLE (57) was used to constrain water. The simulations were initialized in the canonical ensemble with initial velocities taken from a Maxwell-Boltzmann distribution at 300 K. The initialization procedure described in (58) was used to start the simulations smoothly. Subsequently, all the atoms were set free and the production simulations were calculated.

Replicates were built by considering the two slightly different conformations of the protein structure (Monomer A and Monomer B) and assigning sets of different starting velocities. In this way, six simulations were set up for each condition, three for each monomer. The simulation time calculated for each replicate was 100 ns, amounting to 600 ns per condition, in a total simulation time of 4.1 μs.

The MD simulations of the E187A and E115A mutants were performed using the same methodology described above. The starting points of these simulations were the final conformations of the wild type simulations, but individually substituting the Glu residues by Ala. The simulations were then reinitialized using the coordinates and velocities of the final step of the wild type runs.

Database accession codes. The atomic coordinates and the structure factor

amplitudes of the Tgl crystal structure were deposited in the PDB under the accession numbers 4PA5 and 4P8I, for the structures of Tgl-cysteamine and Tgl, respectively.

ClustalW alignment of Tgl proteins from selected* Bacillus *and* Clostridium *species and construction of Figure 2.5. The sequence of Tgl orthologues was aligned using the ClustalW sequence alignment program (www.wbi.ac.uk). Sequence Logo (weblogo.berkeley.edu, (59)) was then used to display the residue conservancy of the blocks of sequences that composed both walls of the tunnel. The Tgl orthologues (and according Genebank accession numbers) used were: *Bacillus subtilis* 168 (NP_39100 5.1); *B. thuringiensis* serovar *andalousiensis* BGSC 4AW1 (ZP_04098104.1); *B. anthracis* str. Ames (NP_8464 11.1); *B. cereus* ATCC 4342 (ZP_0428561 7.1); *B. thuringiensis* serovar *tochigiensis* BGSC 4Y1 (ZP_04147231.1); *B. anthracis* CI (YP_003793665.1); *B. cereus* ATCC 14579 (NP_833682.1); *B. mycoides* DSM 2048 (ZP_04170313.1); *B. mycoides* Rock3-17 (ZP_0415833 0.1); *B. pseudomycoides* DSM 12442 (ZP_0415262 5.1); *B. cereus* Rock3-44 (ZP_042187001); *B. megaterium* DSM 319 (YP_003597620.1); *B. megaterium* QM B1551 (YP_003562903.1); *Clostridium botulinum* Ba4 str. 657 (YP_002863365.1); *Clostridium* botulinum F str. Langeland (YP_0013915 94.1); *C. carboxidivorans* P7 (ZP_05395083.1); *C. kluyveri* DSM 555 (YP_0013 93940.1); *Brevibacillus brevis* NBRC 100599 (YP_002770342.1); *Paenibacillus polymyxa* E681 (YP_003871153.1); *P. polymyxa* SC2 (YP_003947364.1); *B. atrophaeus* 1942 (YP_003974545.1); *B. amyloliquefaciens* DSM7 (YP_00392 1520.1); *B. amyloliquefaciens* FZB42 (YP_001422395.1); *B. licheniformis* ATCC 14580 (YP_080406.1); *B. pumilus* ATCC 7061 (ZP_03053 933.1); *B. clausii* KSM-K16 (YP_175687.1); *Paenibacillus* sp. oral taxon 786 str. D14 (ZP _04851688.1); *Paenibacillus* sp. JDR-2 (YP_003009106.1); *Bacillus* sp. SG-1 (ZP_01860758.1); *Bacillus* sp. NRRL B-14911 (ZP_01171363.1); *B. halodurans* C-125 (NP_244838.1); *B. pseudofirmus* OF4 (YP_003425810.1); *Bacillus* sp. m3-13 (ZP_07708543.1); *Geobacillus* sp. C56-T3 (YP_003670 219.1); *G. kaustophilus* HTA426 (YP_148767.1); *Geobacillus* sp. Y412MC61 (YP_003254034.1); *G. thermodenitrificans* NG80-2 (YP_001126953.1); *G. thermoglucosidasius* C56-YS93 (ZP_0681129 5.1); *Bacillus* sp. NRRL B-14911 (ZP_01172648.1); *Bacillus* sp. 2-A-57-CT2 (EFV79150.1).

RESULTS

Structure of Tgl. The two refined X-ray structures of Tgl were obtained from crystals soaked in a solution of cystamine, a TGase inhibitor (60). The X-ray structures of Tgl, in apo-form and bound to cysteamine (the reduced form

Table 2.3. Data collection, phasing and refinement statistics.

	Native	Ga derivative	Apo-form	Cyst[a] complex
Data collection				
Beamline	X11, DESY	X11, DESY	X06SA, SLS	ID14-1, ESRF
Wavelength (Å)	0.905	1.712	0.918	0.934
Resolution[b] (Å)	39.18 – 2.54 (2.55 – 2.54)	36.42 – 3.06 (3.08 – 3.06)	47.85 – 1.85 (1.91 – 1.85)	36.57 – 1.86 (1.93 – 1.86)
Space group	$P4_1$	$P4_1$	$P4_1$	$P4_1$
Unit cell (Å)	98.11, 98.11, 65.13	98.21, 98.21, 65.17	97.94, 97.94, 66.18	98.35, 98.35, 65.81
Total reflections	208 867 (1 192)	51 164 (284)	353 354 (23 750)	205 919 (21 586)
Unique reflections	19 967 (144)	11455 (120)	47 409 (3 422)	44 619 (5 201)
Multiplicity	105 (8.3)	4.5 (2.4)	7.5 (6.9)	4.1 (4.2)
Completeness (%)	96.9 (61.0)	96.6 (48.6)	88.6 (64.7)	84.8 (71.8)
Mean I/σ(I)	41.2 (8.1)	27.4 (4.5)	29.5 (3.9)	12.1 (1.6)
$R_{merge\ (\%)}$	4.5 (26.1)	8.3 (36.3)	3.8 (43.4)	5.9 (88.5)
CC(1/2)	1.0 (0.97)	0.99 (0.66)	0.99 (0.96)	0.99 (0.85)
Refinement				
R_{work}/R_{free}			15.8/20.8	16.7/21.3
No. non-H atoms (per asymmetric unit)				
Protein			3970	3992
Solvent			239	204
Ligands			55	71
Rmsd				
Bond lengths (Å)			0.011	0.011
Bond angles (°)			1.25	1.27
Ramachandran favored (%)			94	95
Ramachandran allowed (%)			5.4	4
Ramachandran outliers (%)			0.6	1
B-factors (Å^2)				
Protein			51.60	52.50
Solvent			44.40	45.70
Ligands			54.90	66.20

[a]Tgl:cysteamine complex; [b]values in parentheses are for highest-resolution shell.

of cystamine), where refined to 1.85 Å and 1.86 Å with R/R_{free} values of 15.8/20.8% and 16.6/21.2%, respectively (relevant statistics are listed in Table 2.3). Both the refined apo-form and Tgl:cysteamine models comprise 2 molecules of 245 amino acid residues each. The structures of apo-Tgl and Tgl:cysteamine complex are very similar (see materials and methods) thus, only the structure obtained for the Tgl:cysteamine crystal will be discussed in detail. The Tgl:cysteamine complex comprises a cysteamine molecule in chain A which is absent from chain B (Figure 2.2). Ramachandran plot shows 95% of the residues in favored regions and 1% in the disallowed region (61). The electron density maps are generally of good quality, except for three disordered regions: residues 1-20 in both chains and residues 180-184 and 219-224 in chain B. Each molecule has approximate dimensions of 60 x 50 x 30 Å and comprises three β-sheets (two-, four- and six-β-stranded) and ten helices, of which three are 3_{10} helices (Figure 2.3).

The catalytic residue Cys116, previously shown to be essential for the activity of Tgl both *in vivo* and *in vitro* (31,62), is located at the N-terminus of a long helix, α6 (arrow in Figure 2.3). The electron density maps suggest a high degree of flexibility around the active site residues and in both monomers, Glu115 appears in two distinct conformations (Figure 2.2). In monomer A and additional electron density consistently appeared around the Sγ. This extra density was modeled as a bounded fragment of cysteamine (-S-CH2-CH2-NH2) with an occupancy of 88% (Figure 2.2). No such density is seen near $Cys116^B$ (Figure 2.2), where the nearby side chains

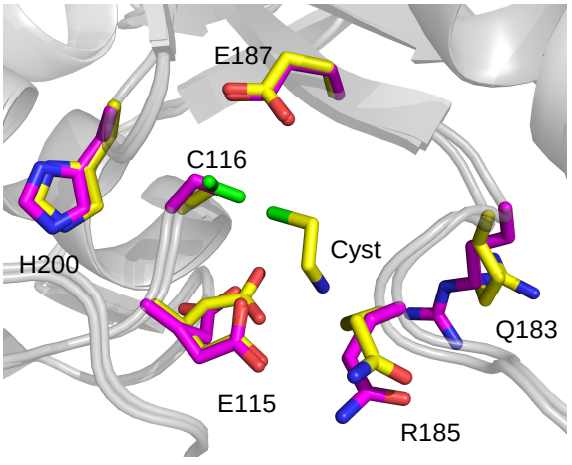


Figure 2.2. Superimposition of the active site region of monomers A and B of the Tgl:cysteamine X-ray structure. Color code: C-, yellow for monomer A, and magenta for monomer B, O- red, N- blue and S- green (the same color code is followed throughout the text for Tgl). Cys116, Gln183 and Arg185 show slightly different conformations in both monomers, while Glu115 displays two different conformations in both monomers. In monomer A, cysteamine (Cyst) is visible.

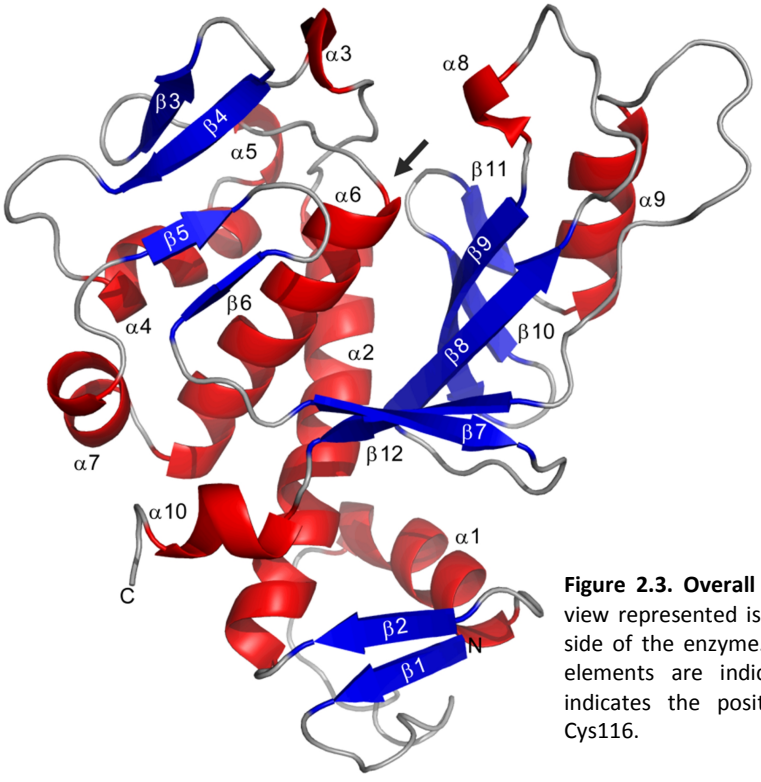


Figure 2.3. Overall structure of Tgl. The view represented is defined as the front-side of the enzyme. Secondary structural elements are indicated and an arrow indicates the position of the catalytic Cys116.

of Gln183^{B} and Arg185^{B} adopt a different conformation from the corresponding residues in chain A (Figure 2.2).

The hydrophobic tunnel of Tgl and substrate accessibility. Strikingly, the structure of Tgl shows a tunnel near the surface that crosses the molecule, from side to side (Figure 2.4A). The tunnel is about 15 Å long (Figure 2.4B, as measured between Trp149 and Ile210). Its narrowest point is located approximately at the center of the tunnel (4.9 Å, measured between Phe69 and Gln183; Figure 2.4B). The front and back entrances of the tunnel are 6.9 Å (from Trp149 to Tyr171) and 7.5 Å (from Trp184 to Phe69) across, respectively (Figure 2.4C,D). The front-side of the tunnel includes residues Phe72, Trp149, Tyr171 and Leu214 (Figure 2.4C). This side of the tunnel presents a network of solvent molecules, including a PEG fragment, while the backside has fewer water molecules (not shown). A conspicuous feature of the backside is the close stacking of Trp184 with Phe217 (nearest distance is ~3.9 Å) (Figure 2.4D). Together with Phe69, these aromatic residues are in parallel packing and create a hydrophobic environment to

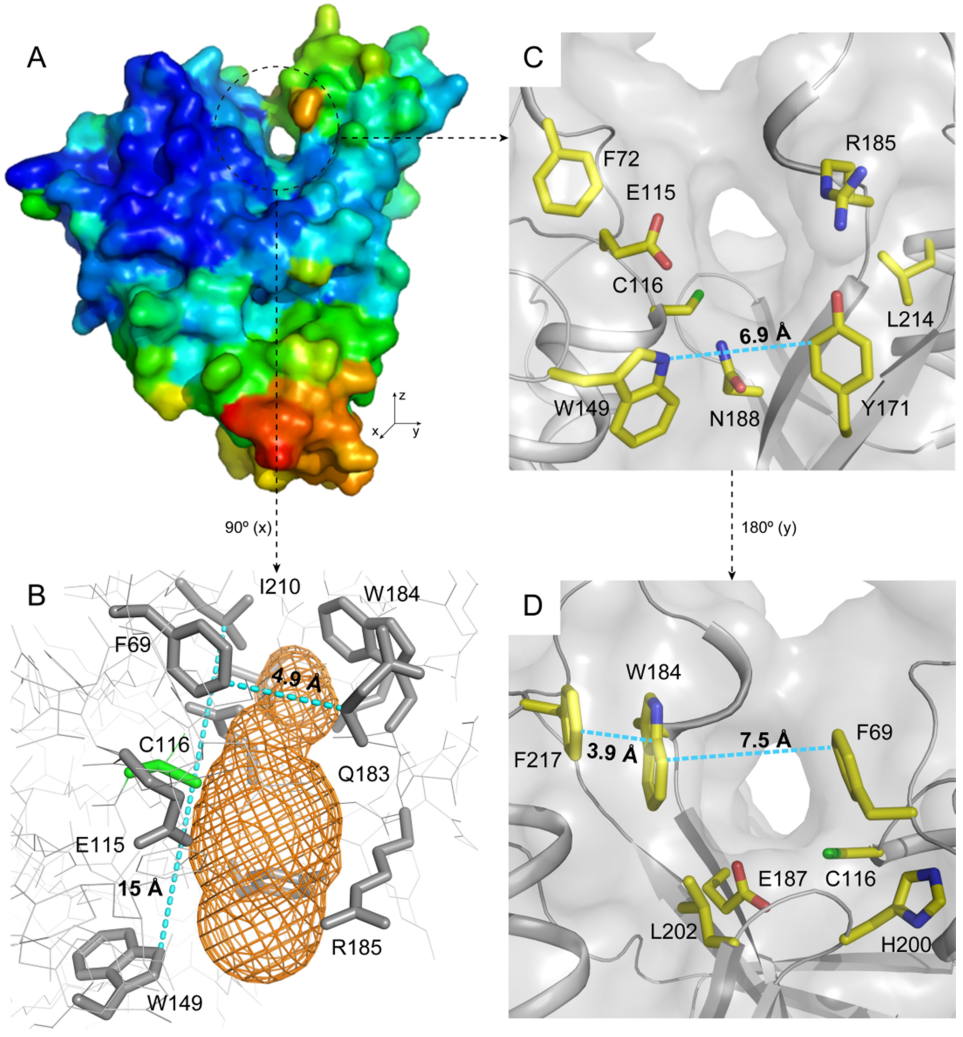


Figure 2.4. The Tgl tunnel. A. Molecular surface representation of Tgl, color-coded according to temperature (B) factors (from blue, lower B factors, to red). The region around the Tgl tunnel is encircled. **B.** Mesh representation of the empty space in the active site region, viewed as a cross section from the top of the Tgl molecule (note that residues forming the ceiling of the tunnel are not represented for clarity). Relevant residues are shown in sticks. **C.** Zoomed view of the front entry of the hydrophobic tunnel with important residues represented in sticks. **D.** Zoomed view of the back entry of the tunnel (180° rotation around the y axis relative to C). In B, C and D, the distances between relevant residues are shown.

which Leu202 also contributes (Figure 2.4D). With the exception of Phe217, all of the aromatic residues at both entrances of the tunnel are highly conserved among Tgl orthologues (Figure 2.5), and may contribute for the insulation of the active site from water.

The left sidewall of the channel, as viewed from its front side entrance,

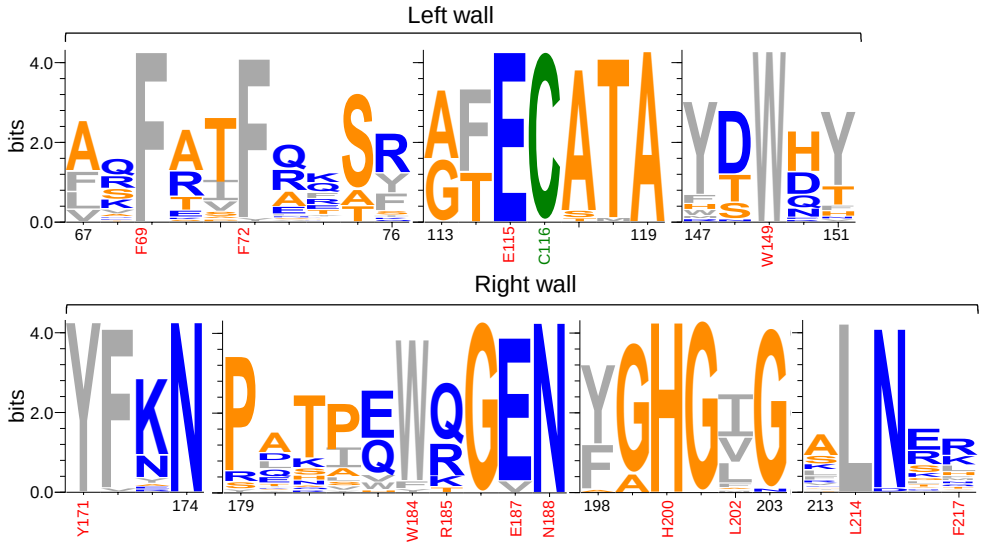


Figure 2.5. Conservation among the tunnel residues. Sequence Logo (59) of the blocks of residues (in single letter notation) that form the left and right wall of the tunnel (numbered according to the Tgl sequence), among 40 selected Tgl orthologues (accession numbers given in materials and methods). Residues are color coded as follows: M, L, I, V, F, W, in grey; A, G, T, P, in orange; E, D, R, K, Q, N, in blue; C, green. Residues represented in panels C and D of Figure 2.4 are highlighted in red.

includes Phe69, Phe72, Glu115, the catalytic Cys116, Ala117, Asp148 and Trp149 (Figure 2.4C,D). The right side comprises Tyr171, Gln183 to Asn188, His200 and Gly201 (Figure 2.4C,D). A sequence logo (59) representation shows a high degree of conservation for most of these residues (Figure 2.5). Some right side residues show higher thermal displacement parameters (B-factors, Figure 2.4A), in particular Arg185, suggesting flexibility. Arg185 also shows some conservation among Tgl orthologues (Figure 2.5). Flexibility at this region of the protein may be important for displacement of the ceiling of the tunnel, allowing the exit of the cross-linked product and the regeneration of the enzyme.

The catalytic core of Tgl adopts the NlpC/P60 papain-like fold. Proteins highly related to Tgl at the primary structure level can only be found in spore-forming bacteria of the *Bacillus* and *Clostridium* genus, and closely related organisms (31,63). However, a DALI search (64) reveals secondary structural similarity between Tgl and several proteins, all of which have in common the presence of an NlpC/P60 domain (Figure 2.6A,B). The highest hits are for the soluble domain (residues 37-162) of the Spr protein, from

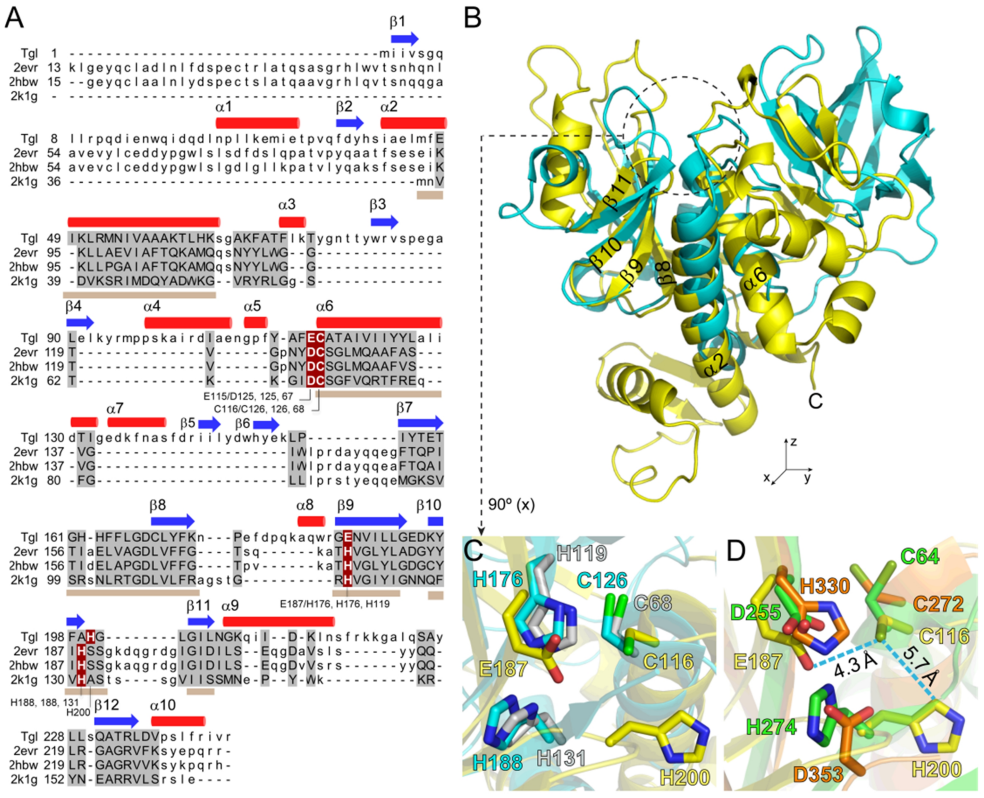


Figure 2.6. The catalytic core of Tgl has an NlpC/P60-like fold. A. Structure-based amino acid sequence alignment of Tgl, NpPCP (pdb code: 2evr), AvPCP (2hbw) and Spr (2kig). The alignment was generated with DALI (64). Helices (red) and β-strands (blue) refer to the Tgl structure. Residues aligned in all four structures are shown in capital letters and highlighted in grey. The secondary structure elements that correspond to the core NlpC/P60 domain are marked with brown lines below the alignment. Actual or presumptive catalytic residues are shown in white against a dark red background, and their numbering shown. **B.** Structural superimposition of Tgl (in yellow) with NpPCP (cyan; other structures are not represented for simplicity). NpPCP (as well as AvPCP) contains an N-terminal SH3-like domain, which is responsible for substrate recognition and is absent in Tgl. The figure highlights the conserved core of the NlpC/P60 domain (corresponding to strands β8-β11 in Tgl, forming a β-sheet packed against α2 and α6; only the clearly visible elements are labeled). The active site region is circled. **C and D.** Superimposition of the putative catalytic residues of Tgl (C-yellow) with: **C**, NpPCP (C- cyan) and Spr (C- gray), and **D**, TGase 3 (C- orange) and MTG (C- green). For animal TGases (e.g., TGase 3), the arrangement of the catalytic residues (Cys-His-Asp) is very similar. Also illustrated is the inversion of the His and Asp/Glu residues of MTG/Tgl when compared to animal TGases. MTG shows a catalytic Cys-Asp dyad instead of the typical triad (26).

Escherichia coli (14), and two homologous cyanobacterial proteins, NpPCP from *Nostoc punctiforme* and AvPCP from *Anabaena variabilis* (13). All three proteins are peptidoglycan cysteine endopeptidases (PCPs) that exhibit Z scores around 7 and RMSD between 2.3-2.7 Å for the 110/101 residues aligned with Tgl. Structural elements shared by Tgl and the PCPs include

helices α2 and α6, which are longer in Tgl, and 4 β-strands (β8 to β11) of an anti-parallel β-sheet element which is packed against α6 (Figure 2.6A,B). A long α-helix with the catalytic Cys residue at its N-terminus (Cys116 in Tgl), against which an anti-parallel β-sheet is packed, is a characteristic of the NlpC/P60 domain (65). Additional catalytic residues are contributed by the β-sheet element. In the papain-like protease group (in which TGases are included), the three-stranded β-sheet of the core is part of a β-barrel with 5 to 8 strands. In the papain-like and in the NlpC/P60 proteases (13,14,65), the β-barrel is formed by 5 strands, whereas in Tgl, this motif is composed by 6 strands (β7-β-12; Figures 2.3 and 2.6A,B). The NlpC/P60 domain is thought to represent an ancestral, minimal fold of the papain-like cysteine proteases and TGases (65). Our observations indicate that the catalytic core of Tgl also adopts the α+β fold of the NlpC/P60 domain. To our knowledge, Tgl is the first structurally characterized TGase in which the ancestral NlpC/P60 domain is clearly recognized. Strikingly, at 28 kDa, Tgl is also the smallest TGase characterized to date.

Active site architecture. NpPCP and AvPCP, and presumably also Spr, make use of a Cys-His-His catalytic triad to cleave the stem peptide of the peptidoglycan molecule (13,14). A closer inspection of the superimposed structures of Tgl, Spr, and NpPCP, in the vicinity of the active site Cys, shows that the position occupied by one of the catalytic His residues in the cell wall endopeptidases (H119 in Spr and H176 in NpPCP) corresponds to Glu187 in Tgl, while the second His (His131 or His188) is close to His200 of Tgl (Figure 2.6C). Both His200 (on a loop between strands β10 and β11) and Glu187 (located on β9), are located within the six-stranded β-sheet element of Tgl (Figures 2.3 and 2.6A,B), lending support to the notion that both could be catalytic residues (but see below).

We then wanted to inspect whether the region around Cys116 of Tgl could also be superimposed to the catalytic residues of MTG, the only other current structural representative of a microbial TGase, and of TGase 3, an archetypal animal TGase. In TGases, catalysis involves a triad consisting of Cys-His-Asp or a Cys-Asp diad (for MTG) (Figure 2.1). The relative position of the catalytic residues in MTG and TGase 3 results from a permutation, such

that the His and Asp residues in MTG correspond to the Asp and His residues in TGase 3, respectively (Figure 2.6D) (26). This is also a functional permutation as Asp255 of MTG plays the role of the catalytic His of animal TGases, while His274 of MTG is not essential for catalysis (26). The superimposition of the catalytic Cys in Tgl (Cys116), MTG (Cys64) and TGase 3 (Cys272), places His200 in the same relative position of His274 in MTG, raising the possibility that His200 is also not essential for catalysis. Furthermore, it also suggests that Glu187 in Tgl could be a catalytic residue. Not only it occupies the position equivalent to His330 in TGase 3 or Asp255 in MTG, but the nearest side-chain atoms of Glu187 are close (~4.3 Å) to the Sγ atom of the catalytic Cys116 residue (Figure 2.6D).

Probing the active site of Tgl. In an attempt to experimentally define the catalytic residues of Tgl, we purified forms of the enzyme bearing single Ala substitutions of residues Cys116 (C116A), Glu187 (E187A), and His200 (H200A) and used them, in parallel with the wild type protein, in two types of activity assays. First, we monitored the incorporation of the fluorescent K donor substrate, dansylcadaverine (dansyl-cd), into bovine serum albumin (BSA, the Q substrate). Tgl^{wt} catalyzed the transfer of dansyl-cd to BSA efficiently, resulting in the fluorescent labeling of the protein (Figure 2.7A). Under our experimental conditions, during the first 45 minutes of the assay, labeling increased linearly with time (Figure 2.7B). Importantly, cross-linking of BSA, auto-cross-linking of Tgl, or labeling of Tgl by dansyl-cd was not significant during the time of the assay (not shown). In confirmation and extension of earlier results (31), we found that the C116A substitution resulted in loss of enzymatic activity (Figure 2.7A). In contrast, labeling of BSA could be detected for Tgl^{H200A} (albeit at low levels) and Tgl^{E187A}, the latter showing about 60% of the wild type enzymatic activity (Figure 2.7A,B).

Second, we assessed the ability of the various forms of Tgl to cross-link a Glutathione S-transferase (GST) fusion protein. GST (about 30 kDa) serves as a substrate for TGases (45,66). The use of a fusion of GST to the first 202 amino acids of coat protein SpoVID (GST-$SpoVID_{202}$, of 50 kDa) (32), increased the mass difference to Tgl (28 kDa), and allowed the levels of enzyme and substrate to be unambiguously monitored on a Coomassie-

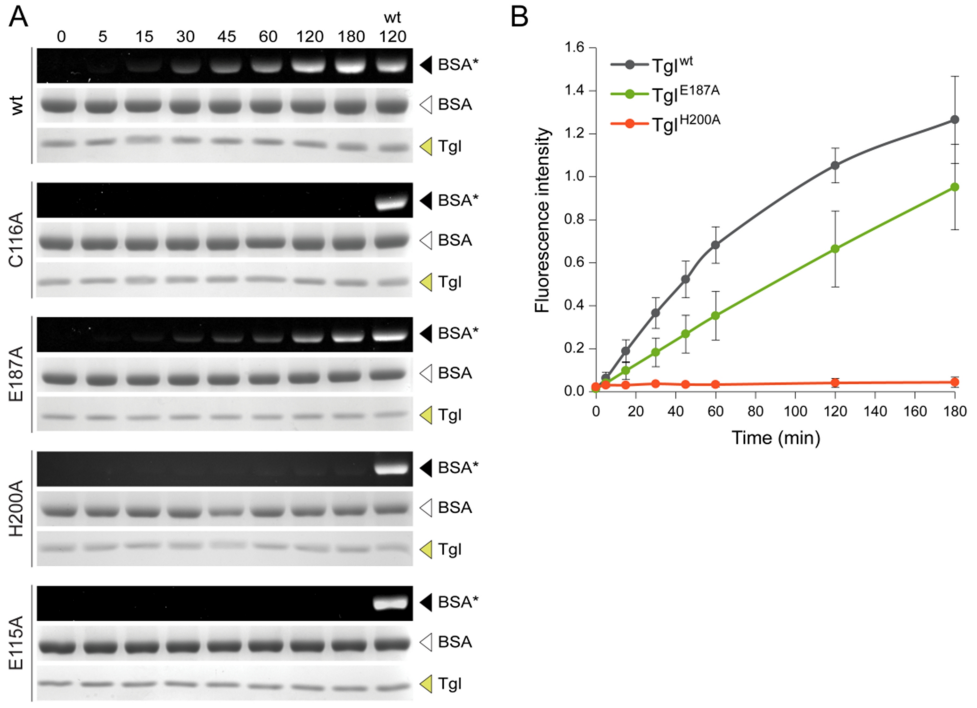


Figure 2.7. Assessment of the enzymatic activity of wild type and mutant forms of Tgl, by BSA labeling. A. Residues Cys116, Glu187, His200 and Glu115 of Tgl were independently substituted by Ala, creating the Tgl^{C116A}, Tgl^{E187A}, Tgl^{H200A} and Tgl^{E115A} mutant enzymes. Enzyme activity was monitored by measuring the incorporation of fluorescent dansylcadaverine into BSA, as detected under UV light (BSA*) after SDS-PAGE analysis of the reaction products. The top rows of each panel represent the fluorescence analysis of samples taken at the indicated reaction times in minutes (numbers above each lane). The middle and bottom rows show the regions of the same gel, stained with Coomassie, where BSA and Tgl migrate, as loading controls. The fluorescence signal was normalized using the level of BSA* detected after 120 min of incubation with Tgl^{wt}. **B.** The normalized fluorescence values were plotted for the various forms of the enzyme, except for Tgl^{C116A} and Tgl^{E115A} for which no fluorescent BSA could be detected; represented is the mean and standard deviation values for each time point. Note that while panel A shows the results of representative assays, panel B includes data for at least three independent experiments.

stained gel. As shown in Figure 2.8A, incubation of Tgl^{wt}, Tgl^{E187A} or Tgl^{H200A} with GST-$SpoVID_{202}$ results in the formation of high molecular weight species over time (indicated by the parenthesis in Figure 2.8A), although at reduced levels for Tgl^{H200A}. In contrast, these species were not detected upon incubation of GST-$SpoVID_{202}$ with Tgl^{C116A} (Figure 2.8B) or when Tgl or GST-$SpoVID_{202}$ were incubated alone (Figure 2.8A,B). The more defined 75, 100, and 250 kDa bands detected after 180 min of incubation with Tgl^{wt}, Tgl^{E187A}, and Tgl^{H200A} (Figure 2.8A, asterisks), were shown by mass spectrometry to contain sequences of both Tgl and GST-$SpoVID_{202}$. While Tgl

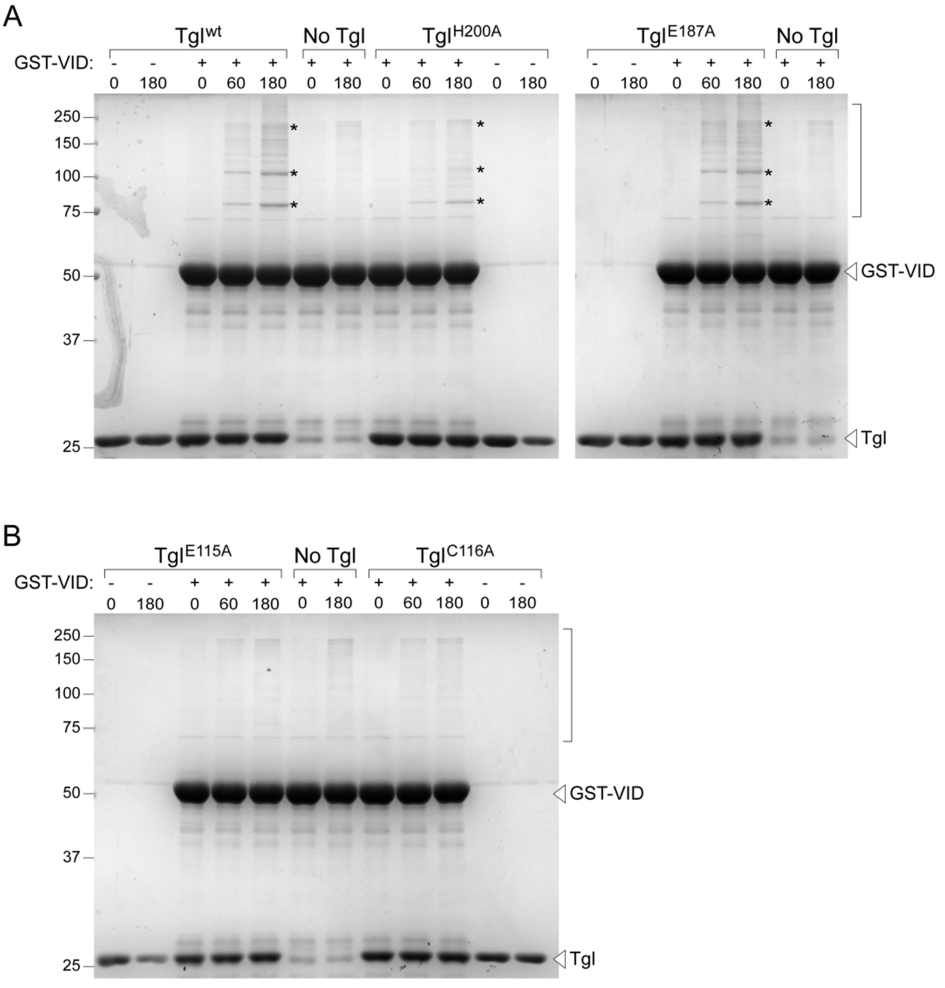


Figure 2.8. Assessment of the enzymatic activity of wild type and mutant forms of TgI, by GST-SpoVID$_{202}$ cross-linking. Crosslinking assays of GST-SpoVID$_{202}$ by TgI^{wt}, TgI^{H200A} or TgI^{E187A} (**A**) and TgI^{E115A} or TgI^{C116A} (**B**). GST-SpoVID$_{202}$ (identified as GST-VID) was incubated with the different forms of TgI for the indicated times (in min; numbers above each lane), after which the reaction mixture was resolved by SDS-PAGE, and the gels stained with Coomassie. The parenthesis indicates the region of the gel where cross-linked products are found. These products are absent from control reactions where GST-SpoVID$_{202}$, TgI^{wt} or the mutant enzymes were incubated alone. The bands labeled with an asterisk were subjected to mass spectrometry and N-terminal sequence analysis. The slight decrease in the level of TgI following 180 min of incubation, especially in the case of TgI^{H200A} or TgI^{E115A}, is likely due to precipitation as it proved to be independent of enzymatic activity, and not seen in all the assays.

alone shows little auto-cross-linking activity under the conditions of the assay (Figure 2.8A, lanes with no GST-SpoVID$_{202}$), these results suggest that TgI (TgI^{wt}, TgI^{E187A} or TgI^{H200A}) cross-links itself to GST-SpoVID$_{202}$. In any event, formation of the 75, 100 and 250 kDa species is the result of TgI

activity. Thus both the dansyl-cd incorporation and cross-linking assays suggest that Cys116, but not Glu187 or His200, is essential for catalysis.

Conformation of the putative catalytic residues at pH 7. The crystal structure of Tgl was obtained at pH 4.0, outside the range for optimal enzymatic activity (67) and this was the only condition at which we were able to obtain stable crystals (see materials and methods). It is possible that our crystallization conditions may have caused artifacts in the relative orientation of critical side chains in the active site region, specifically those of Glu187 and His200 which appear as the most likely catalytic residues along with Cys116 (see above) (Figure 2.6C,D). The contribution of His200 for catalysis, in particular, is difficult to evaluate based on the X-ray structure alone (see Figure 2.1 for the reaction mechanism). Not only is the imidazole ring of His200 too far from Sγ of the catalytic Cys116 to allow proton abstraction (closest distance is ~5.7 Å; Figure 2.6D), as its side chain is also oriented away from the active site Cys (Nδ1 is H-bonded to a water molecule). It is possible that a pH alteration would change the electrostatics at the active site and induce conformational changes resulting in the orientation of His200 towards the catalytic Cys116. Thus, it seemed important to investigate the structural and dynamic properties of Tgl at its optimum pH. To this end, we conducted molecular dynamics (MD) simulations at pH 7, within the optimal interval for enzyme activity (30,67), using the crystal structure determined at pH 4 as the starting point.

In a first step we conducted a titration of all the ionizable groups of the protein, using a computational methodology (58), and obtained predicted p*K*a values for Cys116, and for the neighboring His200 and Glu187 residues, of 14.6, 5.1 and 5.4, respectively. The predicted p*K*a for the catalytic Cys116 may seem too high, as in the reaction mechanism of TGases, this residue loses the proton before catalysis is initiated (Figure 2.1a). However, we note that our calculation uses the structure of Tgl and considers the charge distributions for all atoms of the protein. It also considers the dipole created by the long helix, α6, and the positive charge at the beginning of it, where Cys116 is located (Figure 2.3). The high p*K*a predicted for Cys116 from the X-ray structure is most likely due to the negative potential generated by the

acidic residues in its vicinity, Glu115 and Glu187 (Figure 2.2), which favors protonation and increases the p*K*a over the value displayed by a free Cys residue in solution. However, since our calculations consider a rigid protein, we cannot presently rule out that the p*K*a of Cys116 varies with conformational changes. It could, for instance, be significantly altered when a substrate approaches the binding site of Tgl. Unlike Cys116, our results indicate that the protonation state of Glu187 is completely reversed upon a pH increase from 4 to 7, which most likely influences not only its conformation, but also the conformation of other residues in its vicinity, and in particular, that of His200 (Figures 2.2 and 2.6C,D).

For most residues of Tgl, the protonation state assigned at pH 7.0 was clear. However, this was not the case for His200. Our p*K*a calculations indicate that His200 is either fully protonated or has a proton on Nε2, while the other tautomer (with the proton on Nδ1) has only a marginal probability to occur. Therefore, only the two most likely states of this residue (fully protonated and neutral with the proton on Nε2, identified as ImH^+ and Im, respectively in Figure 2.9) were tested. Two sets of long MD simulations were conducted during which the relative position of His200 was assessed by measuring the minimum distance between its imidazole ring and the carboxyl group of Glu187 (Figure 2.9A,C). The results indicate that the behavior of His200 is strikingly different depending on its protonation state. In five out of six replicates, the fully protonated His200 (ImH^+) reaches a distance of less than 3 Å from Glu187 (Figure 2.9A) and this distance corresponds to the formation of a salt bridge. The interaction of His200 and Glu187 is clearly seen in the representation of the average structure of the last 5 ns of simulation (Figure 2.9B, monomers A2, A3, B1 and B3). In sharp contrast, partially deprotonated His200 (Im) maintains the orientation found in the crystal structure (Figure 2.9C,D). To test the reversibility of the movement of the side chain of His200 upon deprotonation, a new set of simulations were conducted. As starting points for the new simulations, we used the five final structures in which fully protonated His200 approached Glu187 (as shown in Figure 2.9A). For monomers A2, A3, and B1 to B3, the His200 residue was deprotonated and the simulations were continued for a further 100 ns (Figure 2.9E,F). The results indicate that upon deprotonation,

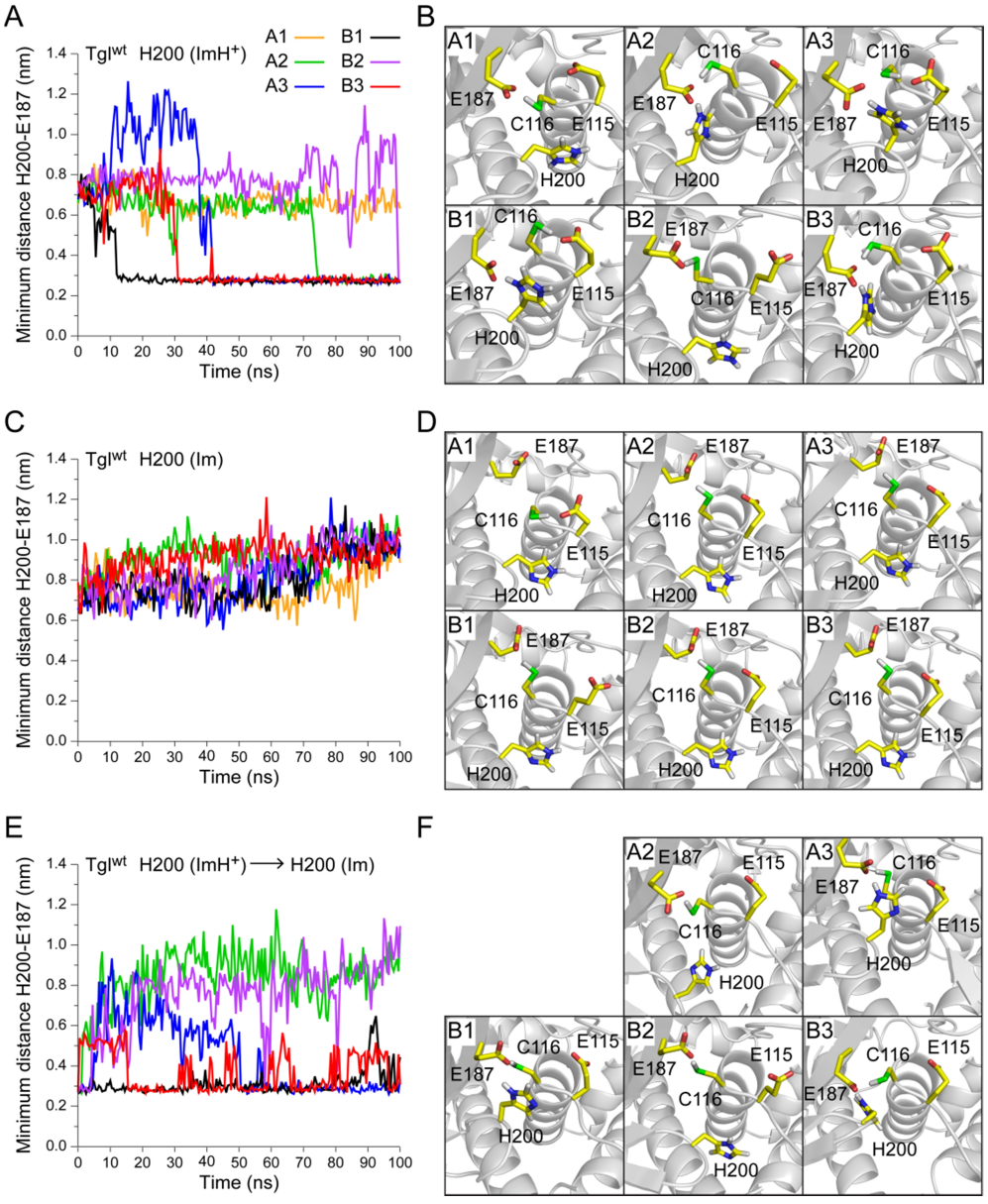


Figure 2.9. Behaviour of His200 during MD simulations. MD simulations were performed with Tgl^{wt} considering different protonation states of His200. Panels A, C and E show the minimum distance between His200 and Glu187, while panels B, D and F show the average structure of the active site of Tgl^{wt}, during the last 5 ns of simulation. Two molecules (monomers A and B) were observed in the final model of the X-ray structure and both were simulated in triplicate (the colour code in A applies to panels C and E). Simulations were performed with Tgl^{wt}, with fully protonated His200 (A, ImH$^+$) or deprotonated His200 on Nδ1 (C, Im). The simulations in panel E used as starting point, the end structures obtained for the simulations with His200 protonated (ImH$^+$, panel A), where His200 approached Glu187 (A2, A3 and B1-B3); His200 was then deprotonated at Nδ1 and a further 100 ns of simulation were conducted. The simulations show that His200 can interact with Glu187 and less frequently with Glu115, but never with Cys116 (see also Tables 2.4 and 2.5).

His200 can form a hydrogen bond with Glu187, albeit less frequently than in the fully protonated state (Figure 2.9A,E). Thus, the results indicate that the conformation of His200 in the crystal structure is likely due to the low pH used during crystallization, and that this residue can change its conformation towards the active site pocket, specifically towards Glu187.

Therefore, both the behavior and protonation states of Glu187 and His200, indicate that the orientation of the putative catalytic residues in the crystal structure reflects, at least in part, the crystallization conditions.

Interactions among residues near the catalytic Cys. A comparison of Tgl to the catalytic core of other TGases and of the NlpC/P60 endopeptidases, placed Glu187 and His200 in the proper position to serve as catalytic partners for Cys116. However, independent Ala substitutions of both residues reduced, but did not eliminate the activity of Tgl. The MD studies described above suggest that the conformation of side chains of residues in the active site region seen in the crystal structure reflects possible artifacts of the crystallization conditions. As such, one possibility that we considered is that other residues in the vicinity of the catalytic Cys116 could partially compensate for the loss of Glu187 or His200. Although these residues do not stand out by analyzing the structure alone they could become apparent in MD simulations at a pH of 7. We started by trying to gain a further insight into the contribution of His200 and Glu187 to the activity of Tgl so that we could generate hypothesis for experimental testing.

Using MD simulations we analyzed the hydrogen bonding interactions among the Cys116, Glu187 and His200 residues, as well as the correlations between all pairs of hydrogen bonds at pH 7.0 (results are displayed in Tables 2.4 and 2.5). The results suggest that Cys116 and His200 do not interact through hydrogen bonding (Figure 2.9B and Table 2.4). Therefore, His200 probably does not act as the proton acceptor for Cys116 (Figure 2.1a). However, Glu187 forms a hydrogen bond with Cys116 with a high frequency (of about 78%) (Figure 2.9B and Table 2.4). Furthermore, while Glu187 may interact with His200, the simultaneous interaction of Glu187 with both His200 and Cys116 appears unfavorable, as the Cys116-Glu187 and His200-Glu187 pairs show a negative correlation (Table 2.5). Therefore,

and in spite of its presumed orientation towards the active site at pH 7.0 as suggested by our MD analysis, His200 does not appear to be a catalytic residue. This is consistent with the observation that Tgl^{H200A} still retains activity in both labeling and cross-linking assays (Figures 2.7 and 2.8A). In contrast, Glu187 does appear to be catalytic residue serving the role as the proton acceptor for Cys116. However, and paradoxically, Tgl^{E187A} still retains considerable labeling and cross-linking activity (Figures 2.7 and 2.8A), raising the possibility that another residue, near the catalytic Cys116, could compensate for the absence of Glu187.

Table 2.4. Frequency of hydrogen bond occurrence, with His200 protonated (ImH^+).

		Hydrogen bond acceptor						
		Tgl^{wt}			Tgl^{E187A}		Tgl^{E115A}	
		Glu115	Cys116	Glu187	Glu115	Cys116	Glu187	Cys116
Hydrogen bond donor	Cys116	20.2%	---	78.3%	82.7%	---	93.2%	---
	His200	3.1%	0.0%	39.4%	1.5%	0.0%	92.8%	0.0%

Table 2.5. Correlations between hydrogen bonds; simulations with Tgl^{wt} and His200 protonated (ImH^+).

			Correlation		
			His200-Glu187	His200-Glu115	Cys116-Glu187
Cys116-Glu115	Replicate	A1	---	---	-0.12
		A2	0.68	-0.69	-0.96
		A3	0.34	-0.34	-0.94
		B1	0.19	-0.20	-0.91
		B2	---	0.00	---
		B3	0.03	-0.05	-0.34
Cys116-Glu187	Replicate	A1	---	---	
		A2	0.00	0.00	
		A3	-0.40	0.41	
		B1	-0.19	0.19	
		B2	---	---	
		B3	---	---	

A new active site residue. Our MD analysis of the interplay between Cys116 and neighboring residues, suggests that a second Glu residue, Glu115, may form a transient hydrogen bond with Cys116, with a frequency of about 20% (Figure 2.9B and Table 2.4). Glu115, which is present in the crystal structure in its protonated form, displayed a p*K*a of 2.1 in our computational titration. As such at pH 7, Glu115 is most likely deprotonated and may work as the proton acceptor for Cys116 (Figure 2.1a). This suggests that Glu115 could also be a catalytic residue. Furthermore, we found a strong negative correlation between the Cys116-Glu187 and the Cys116-Glu115 pairs (Table 2.5), suggesting that the two acids compete for the catalytic Cys. This could explain why Tgl^{E187A} still shows enzymatic activity (Figures 2.7 and 2.8A). His200 and Glu115 also appear to form a hydrogen bond, but the simulations suggest that this is a rare event (frequency of 3%), especially when compared to formation of the His200-Glu187 salt bridge (Figure 2.9B and Table 2.4; see also above). In addition, the simulations also suggest a negative correlation between the Cys116-Glu115 and His200-Glu115 pairs (Table 2.5). We infer that, as was also seen for Glu187, Glu115 does not tend to simultaneously interact with both the Cys and His residues. Finally, we also found strong positive correlations between the His200-Glu187 and Cys116-Glu115 interactions (Figure 2.9B and Table 2.5). This suggests that the interaction between His200 and Glu187 reduces the Glu187-Cys116 interaction so that Cys116 turns to the free Glu115. Similarly, the Cys116-Glu187 and His200-Glu115 pairs are positively correlated (Table 2.5). It follows that the His200-Glu115 interaction directs the catalytic Cys116 to the alternative Glu187. While strengthening the view that His200 is not a catalytic residue, these observations suggest that Tgl uses a catalytic diad formed by either Cys116-Glu187 or Glu115.

A partially redundant catalytic diad: non-reciprocal substitution of Glu187 by Glu115. The results presented in the preceding section suggested that Tgl has a catalytic diad. This diad is however atypical, as two acidic residues (Glu187 and Glu115) appear redundant. If so, then an E115A substitution would not be expected to eliminate enzymatic activity. To test this prediction, Tgl^{E115A} was overproduced, purified and tested for enzymatic activity through dansyl-cd labeling or protein cross-linking, as described

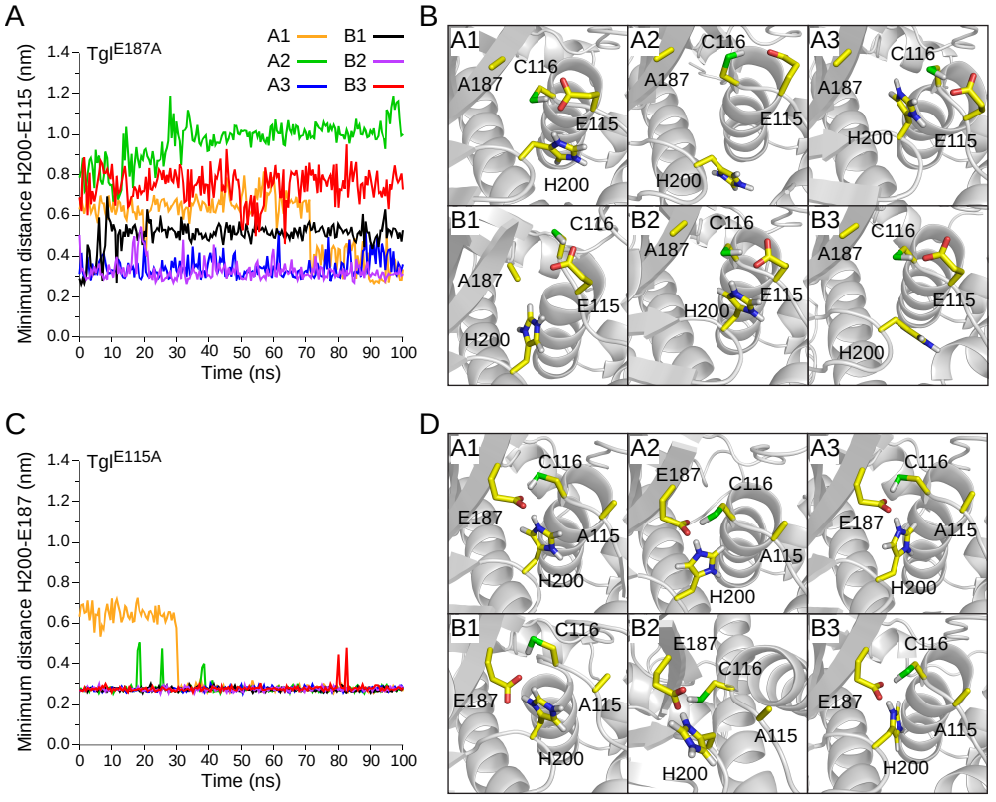


Figure 2.10. Behavior of active site residues during MD simulations with Tgl^{E187A} and Tgl^{E115A}. MD simulations were conducted with *in silico* Tgl^{E187A} (**A and B**) and Tgl^{E115A} (**C and D**) mutants. The MD simulations were performed using as starting points the final conformations of the Tgl^{wt} simulations where His200 is fully protonated (ImH^+, Figure 2.9A), but substituting the Glu residues by Ala. The simulations were conducted for 100 ns and the minimum distance between His200 and Glu187 (A) or Glu115 (C) was measured during the course of the simulations (the colour code in A applies to C). In B and D the average structure of active site residues during the last 5 ns of simulation is represented. The results indicate that the His200-Glu187 interaction becomes stronger in the Tgl^{E115A} than in the wild type enzyme (see also Figure 2.9A,B). Furthermore, it is also clear that in Tgl^{E187A} the simultaneous interaction of Glu115 with both Cys116 and His200 is unfavorable as also seen for Tgl^{wt}. In contrast, for Tgl^{E115A}, Glu187 tends to simultaneously interact with both Cys116 and His200. For the complete analysis of the interactions among the studied residues, refer to Table 2.4.

above. Contrary to our expectations, however, the single E115A substitution rendered Tgl unable to detectably label BSA (Figure 2.7A) or to cross-link GST-SpoVID$_{202}$ (Figure 2.8B). Thus, while Glu115 seems to replace Glu187, the reverse is not observed. To gain insight into this effect, we conducted MD simulations using *in silico* mutants (Figure 2.10). The results suggest that in Tgl^{E187A} the hydrogen bond between Cys116 and Glu115 becomes very likely, occurring with a frequency of about 80%, i.e., 4 times higher than in the wild type enzyme (Table 2.4 and, Figures 2.9B and 2.10B). This is in line

with the idea that in the absence of the Glu187 side chain, Cys116 readily interacts with Glu115. On the other hand, the frequency of the interaction between His200 and Glu115 is low (about 2%; Table 2.4, Figure 2.10A,B). Also, the average structures of the last 5 ns of the simulations with Tgl^{E187A} show that His200 does not tend to interact with the other active site residues (Figure 2.10B). Likewise, the simulations with the Tgl^{E115A} mutant (Figure 2.10C,D), indicate that Cys116 tends to interact with the remaining acidic residue, with the frequency of hydrogen bonding between Cys116 and Glu187 increasing relative to the wild type (from 78.3% to 93.2; Table 2.4, Figures 2.9B and 2.10D). Importantly however, and unlike the results for the wild type enzyme or for Tgl^{E187A}, the absence of Glu115 additionally leads to the very frequent interaction between His200 and Glu187 (frequency of 93%; Table 2.4, Figure 2.10C,D). This suggests that under these conditions the Glu187-His200 interaction is very stable. The once negative correlation between the Cys116-Glu187 and His200-Glu187 (Table 2.5) pairs is now a positive one, and both Cys116 and His200 appear to interact simultaneously with Glu187, which acts as a proton acceptor for both residues (Figure 2.10D, Table 2.4). Considering that in Tgl^{E115A}, Glu187 is the proton acceptor for Cys116 then, the simultaneous interaction with His200 would decrease its ability for proton abstraction and the production of a thiolate ion (Figure 2.1a). The result would be the decreased reactivity of Cys116, possibly explaining the strongly reduced activity of Tgl^{E115A} and the partial redundancy of the Tgl Cys116-Glu187 or Cys116-Glu115 catalytic diad.

DISCUSSION

Here, we present the crystal structure and functional characterization of the bacterial TGase, Tgl. An important finding of our study is that the catalytic core of Tgl is closer to the NlpC/P60 core domain than any other known TGase (65). This domain, likely represents the minimal ancestral unit of the thiol-protease fold (65). *tgl* is confined to the genomes of spore-forming species of the *Bacillus* and *Clostridium* genus, which shared a common ancestral some 3 billion years ago (68). Thus, Tgl is an ancient enzyme.

Interestingly, based on sequence similarity a second putative TGase is detected in *Bacillus subtilis* which appears as an homologue of the animal enzymes (24). Since Tgl homologues are only found in endosporeforming bacteria, it seems possible that this enzyme emerged at the basis of the Firmicutes phylum, and has a sporulation-specific function. That the catalytic core of Tgl adopts the NlpC/P60 fold lends support to the view that TGases have evolved from ancestral proteases (24,65), and also suggests that the adoption of the NlpC/P60 fold by TGases was an early event. In contrast, MTG, the only other microbial TGase that has been structurally characterized, does not share structural similarity with Tgl or with the animal enzymes and may represent a separate evolutionary solution for isopeptide bond formation (26).

A second important finding of this study is that the catalytic center of Tgl is located within a tunnel that crosses the molecule from side to side. The Tgl tunnel is 15 Å-long with both entrances surrounded by hydrophobic residues (Figure 2.4), and it appears wide enough (6 Å) to accommodate a peptide chain backbone bearing a reactive Gln or Lys residue. The higher B-factors, for residues forming the right wall of the Tgl tunnel, appears compatible with its opening following catalysis. Presumably, the Tgl tunnel contributes to waterproofing the active site favoring transamidation over deamidation reactions. The insulation of the catalytic Cys within a hydrophobic tunnel that channels the Q and K substrates to the active site has been proposed as an essential structural requirement for protein cross-linking by TGases (12). Such a tunnel is seen in the active forms of human TGase 2 and 3 and Factor XIIIa (20-23) and may be a conserved feature of the animal enzymes (18). The finding of a tunnel in Tgl, suggests that the adoption of this structural feature was also an early event (12). However, in some enzymes, including MTG, the active site is close to the surface of the molecule (26,69). Therefore, an hydrophobic tunnel containing the catalytic Cys does not seem to be a universal feature of TGases, as suggested previously (12).

Because the tunnel traverses the molecule from side to side and encases the active site, both its entries are most likely docking sites for the Q and K substrates of Tgl. Interestingly while many of the residues at both entrances

of the tunnel are conserved among different Tgl orthologues, some stretches present a high variability (Figure 2.5). It is possible that this variability represents strain specificity towards different substrates. We do not presently know which sides of the tunnel are engaged by the Q and K substrates. Waterproofing of the K acceptor side is important to reduce the potential for hydrolytic reactions, as shown for human TGase 2 (20). The K acceptor side of Tgl may correspond to the more hydrophobic "back" entrance of the tunnel (Figure 2.4D). His200, which as our results suggest, is not a catalytic residue, and is solvent exposed at the "back" entrance of the Tgl tunnel, may be involved in substrate recognition, contributing in that way to the overall activity of the enzyme. An Ala substitution of His274 in MTG, which occupies the same relative position of His200 in Tgl (Figure 2.6D), also reduces, but does not eliminate the activity of the enzyme (24). His274 of MTG may also be involved in substrate interactions (24).

Tgl is only known to cross-link four of the over 70 polypeptides that compose the *B. subtilis* spore coat (31,70-72), and thus appears to be highly selective. Also hinting at a high degree of specificity, Tgl may only recognize proteolytically processed forms of its substrates (72) and unlike other TGases (69), it can be produced to high levels in *E. coli* without causing cell death (30). Both the residues at the tunnel entrances as well as the physical characteristics of the tunnel itself are likely to be important determinants of enzyme specificity.

Tgl has a distinctive catalytic center. In Tgl, Glu187 and His200 are found in the vicinity of the catalytic Cys116, and are superimposable to catalytic residues in other TGases (Figure 2.6D). However, the simultaneous interaction among Cys116, Glu187 and His200 (Table 2.5) are unfavorable, and our MD simulation results suggest that the latter residue cannot serve as the proton acceptor for Cys116 (Figure 2.1a). Together with the experimental results, this indicates that Tgl has a catalytic diad. Why the two simpler (bacterial) TGases employ a catalytic diad (26), whereas the more complex animal enzymes use a triad is presently unclear. Furthermore, our analysis uncovered a peculiar feature of the Tgl active site. While Glu187 appears to be the primary proton acceptor for Cys116, a second acidic residue, in the close vicinity of the catalytic Cys116, can substitute for

Glu187, at least when the latter is substituted by an Ala. The reciprocal substitution (E115A) abrogates enzyme activity, possibly because in the mutant, as suggested by the MD simulations, Glu187 interacts with equal frequency with both Cys116 and His200, reducing its capacity to deprotonate Cys116. Interestingly, Glu115 (but not Glu187) is strictly conserved among Tgl orthologues (Figure 2.5). However, further studies are needed to elucidate its role in catalysis. The partial redundancy of the catalytic diad (Cys116-Glu187/Glu115) implies that catalysis may take place from one or two distinct steric positions. Perhaps different substrates are handled in distinct ways. In any event, this feature of Tgl resembles that of papain, where catalysis occurs as a two-state mechanism via two triads involving two constant residues, Cys25 and His159, and either Asp158 or Asn175 (73). Finally our MD simulations also show that unlike what has been proposed for animal TGases, where the catalytic Cys appears as a thiolate ion in the initial step of the reaction mechanism (Figure 2.1), in Tgl the Cys appears protonated even at very high pH values. It is thus possible that in Tgl the presence of a substrate will induce modifications in the p*K*a of the catalytic Cys, and the formation of the thiolate ion will be concomitant with the initial nucleophilic attack on the Q substrate and the transfer of the proton to Glu187 or Glu115.

Tgl may approach the necessary and sufficient features for protein cross-linking. The ability of this enzyme to function *in vitro* under simple reaction conditions, together with its trimmed structural features, high yield production in active form and high specificity, make it a promising candidate for applications in protein labeling and cross-linking. Moreover, the high specificity and Ca^{2+}-independent activity of Tgl also open the way for *in vivo* applications which have so far been unfeasible with the animal enzymes (8).

REFERENCES

1. **Beninati S, Piacentini M.** 2004. The transglutaminase family: an overview: minireview article. Amino acids **26:**367-372.

2. **Griffin M, Casadio R, Bergamini CM.** 2002. Transglutaminases: nature's biological glues. The Biochemical journal **368:**377-396.

3. **Lorand L, Graham RM.** 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature reviews **4:**140-156.

4. **Lorand L.** 2005. Factor XIII and the clotting of fibrinogen: from basic research to medicine. J Thromb Haemost **3:**1337-1348.

5. **Parameswaran KN, Cheng XF, Chen EC, Velasco PT, Wilson JH, Lorand L.** 1997. Hydrolysis of gamma:epsilon isopeptides by cytosolic transglutaminases and by coagulation factor XIIIa. The Journal of biological chemistry **272:**10311-10317.

6. **Bozzini S, Giuliano L, Altomare L, Petrini P, Bandiera A, Conconi MT, Fare S, Tanzi MC.** Enzymatic cross-linking of human recombinant elastin (HELP) as biomimetic approach in vascular tissue engineering. Journal of materials science **22:**2641-2650.

7. **Mero A, Spolaore B, Veronese FM, Fontana A.** 2009. Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. Bioconjugate chemistry **20:**384-389.

8. **Lin CW, Ting AY.** 2006. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins *in vitro* and on the surface of living cells. Journal of the American Chemical Society **128:**4542-4543.

9. **Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE.** 1994. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Sci **3:**1131-1135.

10. **Ahvazi B, Steinert PM.** 2003. A model for the reaction mechanism of the transglutaminase 3 enzyme. Experimental & molecular medicine **35:**228-242.

11. **Boeshans KM, Mueser TC, Ahvazi B.** 2007. A three-dimensional model of the human transglutaminase 1: insights into the understanding of lamellar ichthyosis. Journal of molecular modeling **13:**233-246.

12. **Nemes Z, Petrovski G, Csosz E, Fesus L.** 2005. Structure-function relationships of transglutaminases--a contemporary view. Progress in experimental tumor research. Fortschritte der experimentellen Tumorforschung **38:**19-36.

13. **Xu Q, Sudek S, McMullan D, Miller MD, Geierstanger B, Jones DH, Krishna SS, Spraggon G, Bursalay B, Abdubek P, Acosta C, Ambing E, Astakhova T, Axelrod HL, Carlton D, Caruthers J, Chiu HJ, Clayton T, Deller MC, Duan L, Elias Y, Elsliger MA, Feuerhelm J, Grzechnik SK, Hale J, Won Han G, Haugen J, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Kumar A, Marciano D, Morse AT, Nigoghossian E, Okach L, Oommachen S, Paulsen J, Reyes R, Rife CL, Trout CV, van den Bedem H, Weekes D, White A, Wolf G, Zubieta C, Hodgson KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA.** 2009. Structural basis of murein peptide specificity of a gamma-D-glutamyl-l-diamino acid endopeptidase. Structure **17:**303-313.

14. **Aramini JM, Rossi P, Huang YJ, Zhao L, Jiang M, Maglaqui M, Xiao R, Locke J, Nair R, Rost B, Acton TB, Inouye M, Montelione GT.** 2008. Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from *Escherichia coli*: structural evidence for a novel cysteine peptidase catalytic triad. Biochemistry **47:**9715-9717.

15. **Gundemir S, Colak G, Tucholski J, Johnson GVW.** 2012. Transglutaminase 2: A molecular Swiss army knife. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1823:**406-419.

16. **Klöck C, Khosla C.** 2012. Regulation of the activities of the mammalian transglutaminase family of enzymes. Protein Science **21:**1781-1791.

17. **Kim SY, Jeitner TM, Steinert PM.** 2002. Transglutaminases in disease. Neurochemistry international **40:**85-103.

18. **Iismaa SE, Mearns BM, Lorand L, Graham RM.** 2009. Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. Physiological reviews **89:**991-1023.

19. **Park D, Choi SS, Ha K-S.** 2010. Transglutaminase 2: a multi-functional protein in multiple subcellular compartments. Amino acids **39:**619-631.

20. **Pinkas DM, Strop P, Brunger AT, Khosla C.** 2007. Transglutaminase 2 undergoes a large conformational change upon activation. PLoS biology **5:**e327.

21. **Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM.** 2002. Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. The EMBO journal **21:**2055-2067.

22. **Stieler M, Weber J, Hils M, Kolb P, Heine A, Büchold C, Pasternack R, Klebe G.** 2013. Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. Angewandte Chemie International Edition **52:**11930-11934.

23. **Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM.** 2003. Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. The Journal of biological chemistry **278:**23834-23841.

24. **Makarova KS, Aravind L, Koonin EV.** 1999. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. Protein Sci **8:**1714-1719.

25. **Reiss K, Kirchner E, Gijzen M, Zocher G, Loffelhardt B, Nurnberger T, Stehle T, Brunner F.** 2011. Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria. Journal of Biological Chemistry **286:**42585-42593.

26. **Kashiwagi T, Yokoyama K, Ishikawa K, Ono K, Ejima D, Matsui H, Suzuki E.** 2002. Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. The Journal of biological chemistry **277:**44252-44260.

27. **Yang MT, Chang CH, Wang JM, Wu TK, Wang YK, Chang CY, Li TT.** 2010. Crystal Structure and Inhibition Studies of Transglutaminase from *Streptomyces mobaraense*. Journal of Biological Chemistry **286:**7301-7307.

28. **Paredes CJ, Alsaker KV, Papoutsakis ET.** 2005. A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol **3:**969-978.

29. **Henriques AO, Moran CP, Jr.** 2007. Structure, assembly, and function of the spore surface layers. Annual review of microbiology **61:**555-588.

30. **Placido D, Fernandes CG, Isidro A, Carrondo MA, Henriques AO, Archer M.** 2008. Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. Protein expression and purification **59:**1-8.

31. **Zilhao R, Isticato R, Martins LO, Steil L, Volker U, Ricca E, Moran CP, Jr., Henriques AO.** 2005. Assembly and function of a spore coat-associated transglutaminase of *Bacillus subtilis*. Journal of bacteriology **187:**7753-7764.

32. **Costa T, Isidro AL, Moran CP, Jr., Henriques AO.** 2006. Interaction between coat morphogenetic proteins SafA and SpoVID. Journal of bacteriology **188:**7731-7741.

33. **Studier FW.** 2005. Protein production by auto-induction in high density shaking cultures. Protein expression and purification **41:**207-234.

34. **Kabsch W.** 2010. Xds. Acta Crystallogr D Biol Crystallogr **66:**125-132.

35. **Vonrhein C, Blanc E, Roversi P, Bricogne G.** 2007. Automated structure solution with autoSHARP. Methods in molecular biology (Clifton, N.J **364:**215-230.

36. **Sheldrick GM.** 2008. A short history of SHELX. Acta Crystallogr A **64:**112-122.

37. **Abrahams JP, Leslie AG.** 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr **52:**30-42.

38. **Cowtan K.** 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr **62:**1002-1011.

39. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66:**486-501.

40. **Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Smart OS, Vonrhein C, Womack TO.** 2009. BUSTER, version 2.8. 0. Cambridge, United Kingdom: Global Phasing Ltd.

41. **Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD.** 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr **68:**352-367.

42. **Laskowski RA, MacArthur MW, Moss DS, Thornton JM.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. **26:**283-291.

43. **Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC.** 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr **66:**12-21.

44. **DeLano.** 2002. The PyMOL Molecular Graphics System Palo Alto, CA, USA.

45. **Sugimura Y, Hosono M, Wada F, Yoshimura T, Maki M, Hitomi K.** 2006. Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGase 2 and Factor XIIIa. The Journal of biological chemistry **281:**17699-17706.

46. **Rasband WS** 1997-2012, posting date. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/. [Online.]

47. **Moulaei T, Stuchlik O, Reed M, Yuan W, Pohl J, Lu W, Haugh-Krumpe L, O'Keefe BR, Wlodawer A.** 2010. Topology of the disulfide bonds in the antiviral lectin scytovirin. Protein Sci **19:**1649-1661.

48. **Bashford D.** 1997. An object-oriented programming suite for electrostatic effects in biological molecules., p. 233-240. *In* Ishikawa Y, Oldehoeft, R. R., Reynders, J. V. W., Tholburn, M. (ed.), Scientific Computing in Object-Oriented Parallel Environments, vol. 1343. Springer, Berlin.

49. **Bashford D, Gerwert K.** 1992. Electrostatic calculations of the pka values of ionizable groups in bacteriorhodopsin. J. Mol. Biol. **224:**473-486.

50. **Baptista AM, Soares CM.** 2001. Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. J. Phys. Chem. B **105:**293-309.

51. **Teixeira VH, Cunha CA, Machuqueiro M, Oliveira AS, Victor BL, Soares CM, Baptista AM.** 2005. On the use of different dielectric constants for computing individual and pairwise terms in poisson-boltzmann studies of protein ionization equilibrium. The journal of physical chemistry **109:**14691-14706.

52. **H. J. C. Berendsen DvdSaRvD.** 1995. GROMACS: A message-passing parallel molecular dynamics implementation. Comp. Phys. Commun. **91:**43-56.

53. **Hess B, Kutzner, C., van der Spoel, D. & Lindahl, E. .** 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. **4:** 435-447.

54. **Oostenbrink C, Villa A, Mark AE, van Gunsteren WF.** 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. Journal of computational chemistry **25:**1656-1676.

55. **Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG.** 1995. A Smooth Particle Mesh Ewald Method. J. Chem. Phys. **103:**8577-8593.

56. **Hess B, Bekker H, Berendsen HJC, Fraaije JGEM.** 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. **18:**1463-1472.

57. **Miyamoto S, Kollman PA.** 1992. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. **13:**952-962.

58. **Lousa D, Cianci M, Helliwell JR, Halling PJ, Baptista AM, Soares CM.** 2012. Interaction of counterions with subtilisin in acetonitrile: insights from molecular dynamics simulations. The journal of physical chemistry **116:**5838-5848.

59. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome research **14:**1188-1190.

60. **Jeitner TM, Delikatny EJ, Ahlqvist J, Capper H, Cooper AJ.** 2005. Mechanism for the inhibition of transglutaminase 2 by cystamine. Biochemical pharmacology **69:**961-970.

61. **Laskowski RA, MacArthur, M.W., Moss D.S., Thornton, J.M.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography **26:**283-291.

62. **Kobayashi K, Hashiguchi K, Yokozeki K, Yamanaka S.** 1998. Molecular cloning of the transglutaminase gene from *Bacillus subtilis* and its expression in *Escherichia coli*. Bioscience, biotechnology, and biochemistry **62:**1109-1114.

63. **Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ.** 2012. Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. Environmental microbiology **14:**2870-2890.

64. **Holm L, Sander C.** 1993. Protein structure comparison by alignment of distance matrices. J Mol Biol **233:**123-138.

65. **Anantharaman V, Aravind L.** 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome biology **4:**R11.

66. **Piredda L, Farrace MG, Lo Bello M, Malorni W, Melino G, Petruzzelli R, Piacentini M.** 1999. Identification of 'tissue' transglutaminase binding proteins in neural cells committed to apoptosis. Faseb J **13:**355-364.

67. **Kobayashi K, Suzuki SI, Izawa Y, Miwa K, Yamanaka S.** 1998. Transglutaminase in sporulating cells of *Bacillus subtilis*. J Gen Appl Microbiol **44:**85-91.

68. **Battistuzzi FU, Feijao A, Hedges SB.** 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC evolutionary biology **4:**44.

69. **Reiss K, Kirchner E, Gijzen M, Zocher G, Loffelhardt B, Nurnberger T, Stehle T, Brunner F.** Structural and phylogenetic analyses of the GP42 transglutaminase from Phytophthora sojae reveal an evolutionary relationship between oomycetes and marine Vibrio bacteria. The Journal of biological chemistry **286:**42585-42593.

70. **Monroe A, Setlow P.** 2006. Localization of the transglutaminase cross-linking sites in the *Bacillus subtilis* spore coat protein GerQ. Journal of bacteriology **188:**7609-7616.

71. **Ragkousi K, Setlow P.** 2004. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. Journal of bacteriology **186:**5567-5575.

72. **Kuwana R, Okuda N, Takamatsu H, Watabe K.** 2006. Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. Journal of biochemistry **139:**887-901.

73. **Wang J, Xiang YF, Lim C.** 1994. The double catalytic triad, Cys25-His159-Asp158 and Cys25-His159-Asn175, in papain catalysis: role of Asp158 and Asn175. Protein engineering **7:**75-82.

Chapter 3

Auto-regulation of SafA assembly via the subcellular localization of a protein cross-linking enzyme

This Chapter contains data to be published in: Catarina G. Fernandes and Adriano O. Henriques. 2014. *Auto-regulation of SafA assembly via the subcellular localization of a protein cross-linking enzyme*. To be submitted to The Journal of Bacteriology.

The author of this dissertation performed all the experiments described in this chapter with the exception of the immunoelectron microscopy experiments which were conducted at the Core Microscopy Unit, Emory University, USA, under the supervision of Hong Yi; also, strains AH2850, AH4745, and AH5074 were constructed by Teresa Costa, Fátima Pereira and, Filipa Nunes, respectively (Microbial Development laboratory).

ABSTRACT

The coat of *Bacillus subtilis* spores is a multiprotein structure that protects the underlying spore cortex peptidoglycan layer against the action of lytic enzymes, and arbitrates many of the environmental interactions of the spore. The coat is differentiated into two main sub-structures, an inner and an outer layer. Morphogenetic protein SafA governs assembly of the inner coat, whereas CotE is required for outer coat assembly. The *safA* gene codes for a full-length protein of 43 kDa (SafA) and additionally, for proteins of 21 (N21) and 30 kDa (C30) corresponding to the N- and C-terminal moieties of SafA. SafA localizes to the cortex/coat interface and probably helps in the connection of these two structures of the spore. Both SafA and C30 are substrates for Tgl, a coat-associated transglutaminase that cross-links proteins through ε-(γ-glutamyl)lysyl isopeptide bonds. We now show that, in wild type spores, SafA and C30 are distributed between the inner coat and cortex layers and are present in a form that resists extraction. In contrast, deletion of *tgl* greatly increases the extractability of SafA, which is mainly associated with the cortex layer. Tgl, itself, is mostly located in the cortex and inner coat layers. The localization of Tgl-CFP around the developing spore is strongly, but not exclusively dependent on *safA,* however, the association of Tgl with the cortex is dependent on *safA*. Together, our results suggest a two-step assembly pathway in which Tgl is first recruited to the forming spore in a manner that is at least partially independent on SafA, and then is drafted to the cortex by the morphogenetic protein. Tgl, in turn, promotes the conversion of cortex-associated SafA into a form that resists extraction, possibly by cross-linking it to the cortex and/or to a cortex-associated protein. Therefore, the final assembly state of SafA relies on an auto-regulatory pathway that determines the subcellular localization of a protein cross-linking enzyme.

INTRODUCTION

A group of bacteria in the Firmicutes phylum, including those belonging to the well-known *Bacillus* and *Clostridia* genus have the ability to differentiate highly resistant spores. Sporulation is a remarkable illustration of a develop-

ment process and the construction of the outermost layer of the spore, called the coat, an example of the assembly of a multiprotein, non-membrane bounded, bacterial organelle at a specific subcellular address. For these reasons and because endospores of pathogenic species of *Bacillus* and *Clostridia* are infectious agents (1,2), the sporulation process has been extensively studied. Despite this, many of the aspects of the assembly of the coat still remain elusive. This is in part because the coat is composed by more than 70 different proteins (3-6) whose assembly is regulated by several factors: time, place and levels of expression of the different coat genes, protein-protein interactions, and post-translational modifications of the coat proteins (7,8). Here, we were interested in understanding the assembly of SafA, a key protein in spore coat assembly, and its dependency on a transglutaminase, Tgl, an enzyme that cross-links proteins at the spore surface.

The sporulation process in *B. subtilis* takes about 8 hours to be completed and at the end, the mature free spores are composed by three main compartments: the core, which houses a copy of the genome; the cortex, a modified form of peptidoglycan that surrounds the core and confers resistance to heat; and, finally, the coat, that encircles the cortex (7,8). The different coat proteins assemble into two main sub-structures, readily visualized by thin-section electron microscopy: the inner and outer coat layers (3,7). The proper assembly of the different coat proteins affords the spore with resistance to lytic enzymes, noxious chemicals and UV radiation. The coat is also important for the detection of outside signals for germination triggering when conditions are favorable (3,9,10).

During growth, *B. subtilis* divides symmetrically and sporulation initiation is detected by the formation of an asymmetric septum near one of the poles of the cell (11). This leads to the creation of two unequal sized cells, a larger one, denoted mother cell, and a smaller one, called the forespore, which will ultimately give rise to a mature free spore. Once the septum is formed the two septal membranes start to migrate around the forespore, in a process denominated as engulfment. When engulfment is finalized the forespore becomes a free protoplast with a double membrane and fully separated from the surrounding medium (11,12). Soon after, the cortex

starts being assembled and phase dark spores inside the mother cell can be visualized by phase contrast microscopy. This is followed by the appearance of phase bright forespores due to cortex maturation processes. While the assembly of the coat begins soon after asymmetric division, the late stages of its formation take place only following engulfment completion (3,9).

The time and place of expression of each of the genes involved in sporulation is tightly controlled. This is achieved mainly due to four sigma factors of RNA polymerase which are sequentially activated in the forespore or mother cell. Just after asymmetric septum completion σ^F and σ^E become active (in the forespore and mother cell, respectively), while σ^G (forespore) and σ^K (mother cell) will only become active once engulfment is finished (12). The coat proteins are synthesized in the mother cell compartment under the control of either σ^E or σ^K and deposited on the surface of the developing forespore (3).

The proper assembly of the different coat components is largely dependent on a small subset of so-called morphogenetic proteins (3,9,13), which are expressed early in sporulation under the control of σ^E. These proteins are thought to create a scaffold essential for the recruitment of the coat proteins. At the base of this scaffold are SpoIVA, SpoVM and SpoVID whose absence severely impairs coat formation. SpoIVA and SpoVM are also necessary for proper cortex formation (14-17). Two other morphogenetic proteins, SafA and CotE, are responsible for inner and outer coat assembly, respectively (18-21).

When the coat layer is properly assembled, about 30% of the spore coat material found in free spores, resists extraction by different methods (22). Three types of covalent cross-linking have been proposed or detected in this insoluble spore coat material: di-tyrosine, di-sulfide and ε-(γ-glutamyl)lysyl bonds (8,13,23). Tgl, a σ^K dependent transglutaminase, catalyzes the formation of the latter type of cross-links (24,25), and deletion of the *tgl* gene affects coat morphology as visualized by electron microscopy (26). Deletion of *tgl* increases the extractability of several coat proteins, which most likely correspond to Tgl substrates (26,27). These proteins have been proposed, or confirmed, to correspond to YeeK, GerQ (*ywdl*), and two forms

of the SafA protein (24,26-28). During sporulation, due to an alternative site of initiation of translation, three different forms of SafA are produced (29): the full length protein (SafA), which has the morphogenetic role in inner coat assembly, and two others, corresponding to the N and C-terminal parts of the protein with 21 and 30 kDa, $SafA_{N21}$ (N21) and $SafA_{C30}$ (C30), respectively (29,30). Only SafA and C30 are substrates of Tgl (27). While four of the substrates of Tgl have been proposed, nothing is known about their binding partners in the cross-linking reaction.

Surprisingly, while the Tgl substrates are thought to be inner coat proteins (31,32), the localization of Tgl-GFP (20,27) during sporulation appears to be strongly dependent on CotE (responsible for outer coat formation). However, the assembly of Tgl also appears to depend on SafA and GerE (26,27). GerE is a transcriptional regulator whose absence leads to spores without an inner coat and with a defective outer coat (33). Together, these observations suggest that Tgl is present in distinct layers of the coat.

Apart from being cross-linked within the spore lattice, three out of the four substrates of Tgl also appear to suffer a second type of post-translational modification, which is dependent on YabG, a protein with *in vitro* protease activity (34). It has been suggested that YabG processes the substrates of Tgl, which then catalyzes the cross-linking reaction. The processing activity of YabG may be a way to regulate the activity of Tgl during sporulation. While the activity of transglutaminases is highly regulated, either by the presence of cofactors, or by production of the enzymes as zymogens (35-37), Tgl appears to be synthesized in active form (Chapter2 and (38)). However, its activity is only detected late in sporulation, long after the enzyme has been synthesized (24,26).

Here, we examined the subcellular localization of SafA, C30, and Tgl within the spore and their interdependency. We found that SafA and C30 are associated with the cortex layer of mature spores. Cortex-associated SafA is detected in a form that is highly resistant to extraction, and this is dependent on the activity of Tgl. Tgl, itself, localizes predominantly to the cortex and inner coat layers of the spore, and while the initial assembly of the enzyme to the forespore is partially *safA*-independent, the presence of

Tgl in the cortex compartment is reliant on *safA*. Overall, our results show that the association of SafA and Tgl within the spore cortex layer is interdependent.

MATERIALS AND METHODS

Bacterial strains, media and general methods. All plasmid constructions were conducted in *Escherichia coli* DH5α. All *B. subtilis* strains used in this study are derivatives of the wild type MB24 strain. Routine growth of *E. coli* and *B. subtilis* strains was done in Luria-Bertani medium supplemented with the appropriate antibiotic when needed. For sporulation induction, strains were grown in Difco sporulation medium (DSM) (39). For plasmid constructions all fragments were amplified with NZYDNAChange polymerase (NZYTech); for site directed mutagenesis and amplification of whole plasmids Phusion high-fidelity DNA polymerase (Finnzymes) was used. Restriction enzymes were purchased from Fermentas. All new plasmids constructed (apart from plasmid derived from sub-cloning strategies) were sequenced to confirm the expected sequence and the absence of undesired mutations. Insertions of genes at the *amyE* locus were confirmed by loss of extracellular amidase activity; insertions in loci other than *amyE* were confirmed by PCR. All strains, plasmids and oligonucleotide primers are listed on Tables 3.1, 3.2 and 3.3, respectively.

safA *in-frame deletion mutant*. A DNA fragment encompassing 522 bp upstream of the *safA* coding region until codon 45, was amplified with primers safA-364D and safA+135R, using genomic DNA from MB24 as template. A second 522 bp fragment from codon 250 of *safA* to downstream region, was amplified with primers safA+748D and safA+1250R. The two fragments were joined by overlapping PCR using primers safA-364D and safA+1250R. The final fragment, of 1002 bp, contains the coding region of *safA* with an in-frame deletion of codons 46 to 249, flanked by the upstream and downstream regions of *safA*. This fragment was digested with *Sal*I and *EcoR*I and cloned in the same sites of pUC18 (New England Biolabs), producing pCF72. The in-frame deletion mutant of *safA* was constructed by congression: MB24 was co-transformed

with pCF72 and genomic DNA from strain AH5074, bearing a chloramphenicol resistance gene, from pDG364 (40), inserted at *amyE*. Chloramphenicol resistant colonies were then screened for impaired germination (18) by the tetrazolium overlay assay (39). A colony with the desired phenotype was then confirmed to have the in-frame *safA* deletion by PCR, and the strain named AH10297. Upon each round of transformation of AH10297 with genomic DNA or plasmids carrying *safA* derivatives for *amyE* insertion, retention of the in-frame deletion was verified by PCR.

safA *integration at the native locus*. Plasmid pTC204 (41), carrying a *safA-gfp*, was digested with *Pfl*MI and *Eco*RI removing *gfp* and part of the coding region of *safA*. This fragment was substituted by a fragment obtained from pTC88 (41) (contains *safA*) with the same restriction enzymes. The final plasmid, pFP2 contains the full coding region of *safA* and this plasmid was used to transform AOB68 (32) (Δ*safA*::*sp*) via a single crossover event (Campbell type recombination) at the *safA* locus, resulting in strain AH4745.

safA *integrations at the* amyE *locus*. *safA.* We started by constructing a vector containing *safA* with its promoter and terminator regions, which have been determined (18,29). Using genomic DNA from MB24, we amplified the *safA* region with primers safA-169D and safA+1248R. The fragment was digested with *Pst*I and *Sal*I and cloned between the same sites of pUC18 (New England Biolabs), creating pCF66. This fragment was then released from pCF66 with *Pst*I and *BamH*I, and inserted into pLitmus29 (New England Biolabs), to yield pCF73. Finally, the *safA* fragment was released from pCF73 with *Xho*I and *BamH*I, and inserted between the same sites of pMLK83 (42), creating pCF75. This plasmid was then used to transfer *safA* to the *amyE* locus of *B. subtilis*.

Δ*C30*. Mutations M161/164A were introduced into the *safA* gene in pCF75 by site directed mutagenesis, using primers safM161/164D and safM161/164R. This created pCF108.

Insertion of* tgl *at the* amyE *locus. *tgl*. A 1928 bp *tgl* fragment was released from pRZ80 (26) with *Sal*I and *Sac*II and inserted in the same sites of pMLK83 (42), creating pCF10. pCF10 was used to transfer the *tgl* gene to *amyE,* creating strain AH4599.

tgl-3N-cfp. *tgl,* including its promoter region, was amplified from pRZ80 (26) with primers pGem-T T7dir and Tgl3NCFPr, generating a 1384 bp *tgl-3N* fragment. *cfp* was amplified from pDR200 (43) with primers 3N-CFPdir and pDR200R, generating a *3N-cfp* fragment. Subsequently, the *tgl-3N* and *3N-cfp* fragments were joined by overlapping PCR using primers pGem-T T7dir and pDR200R. The final fragment containing *tgl-3N-cfp* was digested with *Sal*I and *Sac*II and inserted in the same sites of pMLK83 (42), yielding pCF42.

tgl-hl4-cfp and tgl-fl3-cfp. Based on the method described in (44), plasmid pCF42 was amplified with primers Tgl-HL4-CFP D and Tgl-HL4-CFP R or Tgl-FL3-CFP D and Tgl-FL3-CFP R, creating plasmids pCF63 and pCF64, respectively. The primers contain the sequence of the *hl4,* or *fl3* linker, flanked by the end of the coding region of *tgl* (5' end), and the beginning of the coding region of *cfp* (3' end). The amplification of pCF42 with each pair of primers leads to the exchange of linker *3N* in pCF42 by the *hl4*, or *fl3* linker, and also eliminates the first codon of *cfp*. The amino acid sequence of the linkers are as follows: HL4 (pCF63) – LAEAAAKEAAAKEAAAKEAAAKAA A; FL3 (pCF64) – LGGGGSGGGGSGGGGSAAA.

Insertion of* tgl *at the* thrC *locus. *tgl*. *tgl* was released from pRZ80 (26) with *Xho*I and *Sal*I. This 1869 bp fragment was incubated with the Klenow fragment to generate blunt ends at both extremities. pDG1664 (45) (a *thrC* insertional vector) was digested with *EcoR*I and treated with the Klenow Fragment. The fragment of pRZ80 was cloned in pDG1664, creating plasmid pCF69, in which *tgl* and *thrC* are in opposite orientations. pCF69 was digested with *BamH*I and religated, to remove part of the downstream sequence of *tgl*, producing pCF74, which contains *tgl* flanked by its native promoter and terminator regions.

tgl-hl4-cfp. Primers tgl+703D and cfp+717R tgl+735R were used to amplify the *hl4-cfp* sequence from pCF63. This 847 bp fragment contains *hl4-cfp* flanked by the end of the coding region of *tgl* (5' end) and the region just downstream of it (3' end). The flanking regions have homology for pCF74 and as such, this fragment was then used as a megaprimer (44) to amplify pCF74. This created pCF80 that contains *tgl-hl4-cfp* flanked by the promoter and terminator regions of the *tgl* gene.

Consensus -35 region. Primers tgl-35D and tgl-35R were used to amplify pCF74 generating plasmid pCF87. The amplification of the plasmid leads to the alteration of the -35 region of the promoter of *tgl*, exchanging it from tttca to tcaca (consensus -35 sequence). To create a plasmid with *tgl-hl4-cfp* with an altered -35 region, pCF80 (contains *tgl-hl4-cfp* under the control of the native promoter of *tgl*) was digested with *BamH*I and *EcoR*I. The released 1014 bp fragment containing '*tgl-hl4-cfp*, was inserted between the *BamH*I and *EcoR*I sites of pCF87. This created pCF89 that contains *tgl-hl4-cfp* under the control of the *tgl* promoter with the consensual -35 element.

tgl_{C116A}. First, we constructed a strain where the *thrC* gene was interrupted by a chloramphenicol resistance gene. pMS38 (46) was digested with *SnaB*I and *Hind*III, releasing a fragment that was cloned between the same sites of pDG1664, replacing the erythromycin resistance cassette in pDG1664 by the chloramphenicol resistance determinant. The resulting plasmid, pCF95, was used to transform AH2255 creating strain AH10361, where the *thrC* gene is interrupted by the chloramphenicol resistance gene, while *tgl* is interrupted by the spectinomycin resistance cassette. Then, pCF74 was amplified with primers tgl+288D and tgl+453R, which led to the elimination of base pairs 313 to 594 of the coding region of *tgl*, creating pCF99. Next, 'tgl_{C116A} was amplified from pCF1 using primers tgl F69Adir and tgl+735R and the resulting fragment was digested with *Sfo*I and *EcoR*I and cloned between the same sites of pCF99, creating pCF109. This restores the *tgl* gene, and tgl_{C116A}, in pCF109, is flanked by the native promoter and terminator regions of *tgl*. pCF109 was then used to transform AH10361, and the insertion of tgl_{C116A} at the *thrC* locus was assessed by gain of resistance to erythromycin and loss of resistance to chloramphenicol.

Insertion of* tgl-hl4-cfp *at the* tgl *locus. pCF63 was digested with *Bgl*II and *Sma*I, releasing a fragment that contains '*tgl-hl4-cfp*. This fragment was cloned in the *BamH*I and *EcoR*V sites of pAH256 (47), resulting in pCF83. Campbell recombination with this plasmid, using MB24 as the recipient, produced strain AH10315.

Table 3.1. Bacterial strains used in this study.

Strain	Genotype and phenotype	Origin/Construction
MB24	*trpC2 metC3*, wild type	Laboratory Stock
AH2255	*trpC2 metC3 Δtgl::sp,* Sp^R	(26)
AH2835	*trpC2 metC3 ΔcotE::cm*, Cm^R	(48)
AH4599	*trpC2 metC3 Δtgl::sp ΔamyE::tgl*, Sp^R, Neo^R	AH2255 x pCF10
AH4745	*trpC2 metC3 ΔsafA::sp ΩsafA*, Sp^R, Neo^R	AOB68 x pFP2
AH5074	*trpC2 metC3 ΔamyE::cm*, Cm^R	MB24 x pDG364
AH10170	*trpC2 metC3 Δtgl::sp ΔamyE::tgl-3N-cfp,* Sp^R, Neo^R	AH2255 x pCF42
AH10232	*trpC2 metC3 ΔcotE::cm Δtgl::sp*, Cm^R, Sp^R	AH2835 x AH2255
AH10271	*trpC2 metC3 Δtgl::sp ΔamyE::tgl-hl4-cfp*, Sp^R, Neo^R	AH2255 x pCF63
AH10272	*trpC2 metC3 Δtgl::sp ΔamyE::tgl-fl3-cfp*, Sp^R, Neo^R	AH2255 x pCF64
AH10297	*trpC2 metC3 ΔsafA ΔamyE::cm*, Cm^R	MB24 x pCF72 x AH5074
AH10302	*trpC2 metC3 ΔsafA ΔamyE::safA*, Neo^R	AH10297 x pCF75
AH10307	*trpC2 metC3 Δtgl::sp ΔthrC::tgl*, Sp^R, Erm^R	AH2255 x pCF74
AH10308	*trpC2 metC3 Δtgl::sp ΔthrC::tgl-hl4-cfp*, Sp^R, Erm^R	AH2255 x pCF80
AH10315	*trpC2 metC3 tglΩtgl-hl4-cfp*, Sp^R	MB24 x pCF83
AH10329	*trpC2 metC3 Δtgl::sp ΔthrC::tgl*(-35), Sp^R, Erm^R	AH2255 x pCF87
AH10331	*trpC2 metC3 Δtgl::sp ΔthrC::tgl-hl4-cfp*(-35), Sp^R, Erm^R	AH2255 x pCF89
AH10336	*trpC2 metC3 ΔspoVE::tet Δtgl::sp Δthrc::tgl-hl4-cfp*(-35), Tet^R, Sp^R, Erm^R	AH10331 x JDB1752
AH10341	*trpC2 metC3 ΔcotE::cm Δtgl::sp Δthrc::tgl-hl4-cfp*(-35), Cm^R, Sp^R, Erm^R	AH10232 x pCF89
AH10352	*trpC2 metC3 ΔsafA ΔamyE::cm Δtgl::sp*, Cm^R ,Sp^R	AH10297 x AH2255
AH10357	*trpC2 metC3 ΔsafA ΔamyE::cm Δtgl::sp Δthrc::tgl-hl4-cfp* (-35) Cm^R ,Sp^R, Erm^R	AH10352 x pCF89
AH10361	*trpC2 metC3 Δtgl::sp Δthrc::cm*, Sp^R, Cm^R	AH2255 x pCF95
AH10383	*trpC2 metC3 ΔsafA ΔamyE::safA*$_{M161/164A}$, Neo^R	AH10297 x pCF108
AH10387	*trpC2 metC3 ΔsafA ΔamyE::safA*$_{M161/164A}$ *Δtgl::sp*, Neo^R, Sp^R	AH10352 x pCF108
AH10388	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{C116A}$, Sp^R, Erm^R	AH103961 x pCF109
AOB68	*trpC2 metC3 ΔsafA::sp*, Sp^R	(32)
JDB1752	*trpC2 metC3 ΔspoVE::tet*, Tet^R	(49)

Antibiotic selection stands for: Cm, chloramphenicol; Erm, erythromycin; Neo, neomycin; Sp, spectinomycin; Tet, tetracycline.

Table 3.2. Plasmids used in this study.

Plasmid	Relevant genotype	Origin
pAH256	cloning vector, *sp*	(47)
pCF1	tgl_{C116A}, in pET30a(+) (Novagen)	Chapter 2
pCF10	*tgl,* in pMLK83	This study
pCF42	*tgl-cfp*, in pMLK83	This study
pCF63	*tgl-hl4-cfp*, derived from pCF42	This study
pCF64	*tgl-fl3-cfp*, derived from pCF42	This study
pCF66	*safA*, in pUC18	This study
pCF69	*tgl*, in pDG1664	This study
pCF72	$safA_{\Delta 46\text{-}249}$, in pUC18	This study
pCF73	*safA*, in pLitmus29	This study
pCF74	*tgl*, derived from pCF69	This study
pCF75	*safA*, in pMLK83	This study
pCF80	*tgl-hl4-cfp*, derived from pCF74	This study
pCF83	*'tgl-hl4-cfp*, in pAH256	This study
pCF87	*tgl*(-35), originated from pCF74	This study
pCF89	*tgl-hl4-cfp* (-35), derived from pCF87	This study
pCF95	*thrC::cm*, derived from pDG1664	This study
pCF99	$tgl_{\Delta 313\text{-}594bp}$, derived from pCF74	This study
pCF108	$safA_{M161/164A}$, derived from pCF75	This study
pCF109	tgl_{C116A}, in pDG1664	This study
pDG364	*amyE::cm*	(40)
pDG1664	*thrC::erm*	(45)
pDR200	*cfp*	(43)
pFP2	*safA, neo*	Laboratory stock
pLitmus29	Cloning vector	New England Biolabs
pMLK83	*amyE::neo*	(42)
pRZ80	*tgl,* in pGem-T (Promega)	(26)
pTC88	*safA*, *cm*	(41)
pTC204	*safA-gfp*	(41)
pMS38	*cm*	(46)
pUC18	Cloning vector	New England Biolabs

Table 3.3. Oligonucleotide primers used in this study.

Name	sequence (5' to 3')
pGem-T T7dir	GGCCCGACGTCGCATGCTCCCGGCCGCCATGGC
Tgl3NCFPr	GTAAACAGTTCTTCGCCTTTTGAAACCAT**GTTGTTGTT**GCGGACGATGCG GAAAAGAGACGG
3N-CFPdir	**AACAACAAC**ATGGTTTCAAAAGGCGAAGAACTG
pDR200R	TGTGGTCTAGAGCTAGCCTAGGCTCGAGAAGCT
Tgl-HL4-CFP D	GTCTCTTTTCCGCATCGTCCGC**CTGGCTGAAGCTGCAGCGAAAGAGGCT GCAGCGAAGGAAGCTGCAGCGAAAGAGGCTGCAGCGAAGGCTGCAG CG**GTTTCAAAAGGCGAAGAACTGTT
Tgl-HL4-CFP R	AACAGTTCTTCGCCTTTTGAAAC**CGCTGCAGCCTTCGCTGCAGCCTCTTTC GCTGCAGCTTCCTTCGCTGCAGCCTCTTTCGCTGCAGCTTCAGCCAG**GCG GACGATGCGGAAAAGAGAC
Tgl-FL3-CFP D	GTCTCTTTTCCGCATCGTCCGC**CTGGGCGGAGGCGGATCAGGCGGAGG CGGATCTGGCGGAGGCGGATCAGCTGCGGCA**GTTTCAAAAGGCGAAGA ACTGTT
Tgl-FL3-CFP R	AACAGTTCTTCGCCTTTTGAAAC**TGCCGCAGCTGATCCGCCTCCGCCAGA TCCGCCTCCGCCTGATCCGCCTCCGCCCAG**GCGGACGATGCGGAAAAGA GAC
tgl+703D	CTGGATGTTCCGTCTCTTTTCCGCATCGTCCGC
cfp+717R tgl+735R	GGAAAATAGGCGATGGGGCTTTTTACTTATAAAGTTCGTCCATGCCAAGT GTAATGCCCGCAGC
tgl-35D	GGCTGTCTTCTGCCTTTTTAACT***CA***CATTGCCCAAGCTCTTTGCATATC
tgl-35R	GATATGCAAAGAGCTTGGGCAATG***TG***AGTTAAAAAGGCAGAAGACAGCC
tgl+288D	GCCGCCTTCAAAAGCGATTCGGGACGCCCATGGTCTTGGAATC
tgl+453R	CCGTTTAAGATTCCAAGACCATGGGCGTCCCGAATCGCTTTTG
tgl F69Adir	GCACAAAAGCGGGGCGAAG***GCT***GCCACTTTTTTAAAAACATACGGG
tgl+735R	GCGGACGATGCGGAAAAGAGACGGAACATCCAG
safA-364D	GCAAGTCGACAATCGGGACAGAAATGAATCTTG
safA+135R	CCCGGGTCATACTGAGGCGATACTATTTTCATTCCAGGCATGATTAAGTC
safA+748D	CATGCCTGGAATGAAAATAGTATCGCCTCAGTATGACCCGGGTTATG
safA+1250R	GGGAATTCTAAGCGTGTCAGTTCTCTCCATTTG
safA-169D	CCATTCCTGCAGATGCACTCCTTGCCCTGAATC
safA+1248R	GCATGTTAAGCGTGTCGACTCTCTCCATTTGGC
safM161/164D	CCACAACAGGAGGCT***GCG***AGTAAT***GCT***GAAAATGCAAATTATCC
safM161/164R	CGGATAATTTGCATTTTC***AGC***ATTACT***CGC***AGCCTCCTGTTGTGG

In **bold** is represented the sequence of the linkers inserted; point mutations introduced are represented in ***bold italic***; underlined sequences represent introduced restriction sites.

Purification of spores for analysis of spore coat proteins. Spores were purified from 24 hr DSM cultures by a two-step gradient of 20% and 50% of Gastrografin (Bayer Schering Pharma) as described in (7,50). For the analysis of spore coat proteins, the amount of spores corresponding to an optical density (λ=580 nm) of 2 or 1 (for 10 and 15 well gels, respectively) were boiled for 8 min with extraction buffer containing SDS, DTT, β-mercapto-ethanol and bromophenol blue (7,50). The extracted proteins were resolved by 15% SDS-PAGE. The gels were stained with Coomassie brilliant blue R-250 or transferred to a nitrocellulose membrane for immunoblot analysis (see below). For decoating assays, spores corresponding to an optical density (λ=580 nm) of 4 or 2 (for 10 or 15 well gels, respectively) were boiled for 5 min with extraction buffer (without bromophenol blue) and the samples were then centrifuged for 2 min at 16200 x *g*. The soluble fraction was removed and placed in a new tube and bromophenol blue was added at a final concentration of 0.05%. This corresponds to the coat fraction. The insoluble fraction, corresponding to decoated spores, was washed twice with TBS-T (50 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 8.0) and divided into two equal volume samples that were incubated at 37° C for 2 hr in 50 mM Tris buffer, pH 8.0, with or without lysozyme (2 mg/ml). The samples were then incubated 5 min at 100° C with SDS protein loading buffer and resolved, along with the half of the volume of the previously saved coat fraction, on 15% SDS-PAGE gels. The gels were stained with Coomassie brilliant blue or transferred to nitrocellulose membranes for immunoblot analysis (see below). Unless otherwise stated, for immunoblot analysis of extracted spore coat proteins or decoating assays, half of the volume of sample was applied to the SDS-PAGE gels for transfer to nitrocellulose membranes, when compared to the sample volume for analysis by Coomassie staining.

Mother cell and forespore fractioning. Cultures were grown at 37° C in DSM without antibiotics and samples were taken at the indicated times after the initiation of sporulation. Following the method described in (26), samples were fractioned and separated into total extract, mother cell, and forespore fractions. 50 μg of sample, for total extract and mother cell fraction, or 30 μg of sample for the forespore fraction were applied to 12%

SDS-PAGE gels and transferred to nitrocellulose membranes for immunoblot analysis (see below).

Immunoblot analysis. Immunoblot analysis was conducted according to the instructions for the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific), using milk as blocking agent, and Tween-20 at final concentration of 0.01%. Further details for each specific antibody can be supplied upon request to the authors. Westerns were developed with Clarity Western ECL Substrate (Biorad). The following antibodies were used: anti-Tgl, 1:15000 (26); anti-GFP (laboratory stock), 1:5000; anti-CotA (laboratory stock) and anti-N21 (laboratory stock), 1:1000; and anti-SafA (32), 1:25,000; secondary peroxidase-conjugated antibodies were used at a concentration of 1:100000 with the exception of western blot analysis with anti-N21, where the secondary antibody was used at a concentration of 1:75000.

Heat and lysozyme resistance testes. Heat and lysozyme resistance tests were performed based on the method described in (39). Cultures were grown at 37° C in DSM for 48 hr after inoculation, and the heat or lysozyme treatment was performed for 20 min. The percentage of heat or lysozyme resistance cells corresponds to the cell count following the heat or lysozyme treatment when compared to the total viable cell count.

Fluorescence microscopy. Strains were grown at 37° C in DSM and culture samples (1 ml) were harvested at different times after the initiation of sporulation, washed with phosphate-buffered saline (PBS), resuspended in ~20 µl of PBS and applied to microscopy slides coated with agarose (1.7%). Images were acquired with a Leica DM 6000B microscope coupled to an aniXon+EM camera (Andor Technologies) and controlled by Metamorph software (Meta Imaging series 7.7, Molecular Devices). Phase contrast (PC) and CFP standard filters were used and image analysis was performed with Meta Imaging series 7.7. The wild type (AH10331) strain was analyzed in three different microscopy experiments while the Δ*safA* (AH10357), Δ*cotE* (AH10341) and Δ*spoVE* (AH10336) strains were analyzed by two independent microscopy experiments. For the quantification of the pattern of fluorescence, the number of cells analyzed, for each strain, are as follows:

wt, 624 (T6), 633 (T8), and 554 (T24); Δ*safA*, 152 (T8), 297 (T10), and 211 (T24); Δ*cotE*, 337 (T8), 372 (T10), and 325 (T24); Δ*spoVE*, 239 (T8), and 383 (T10).

For the determination of the forespore/cell fluorescence ratio, two regions were created: one that encircles the forespore and another that surrounds the cell. These regions were created automatically with tools of the Meta Imaging series 7.7 software after the application of different filters to the PC images (exemplified in Figure 3.9D). The automatic determination of forespore and whole cell regions was possible due to the different light refracting properties of the two compartments, one that is bright (the forespore) and the other that is dark (the cell). For this reason, only cells with phase bright spores were analyzed (for images acquired at T10). The regions created in PC images were then transferred to the CFP image and the fluorescence signal of each region was measured. The fluorescence values were corrected by subtracting the average background signal of each image and afterwards, the value of forespore fluorescence was divided by that of the cell, to calculate the forespore/cell fluorescence ratio, (F_{FS}/F_{cell}). For quantification of fluorescence values in free spores, F_S, tools of the Meta Imaging series 7.7 software were used to determine regions encircling free spores in PC images. These regions were transferred to the CFP images; the fluorescence values were quantified and corrected by subtracting the average background signal. The fluorescence values obtained for both, the forespore/cell ratios, and free spores were analyzed with GraphPad Prism 5 (GraphPad Software, Inc) and showed a non-normal distribution. Hence, to normalize the values obtained in different independent experiments, each individual value obtained was divided by the median value of the wild type strain that had been grown and analyzed on the same day. The same procedure was conducted for individual wild type values. The calculation of the normalized values of the ratio of forespore/cell fluorescence or spore fluorescence is summarized in equations 1 and 2, respectively.

(*eq. 1*) $$(F_{FS}/F_{cell})_{norm} = \frac{(F_{FS}/F_{cell})}{\text{median } (F_{FS}/F_{cell})_{wt}}$$

(*eq. 2*) $$(F_S)_{norm} = \frac{F_S}{\text{median } (F_S)_{wt}}$$

Immunoelectron microscopy. Strains were grown in DSM and collected by centrifugation 24 hr after the initiation of sporulation. Samples of the cultures were fixed and analyzed by immunoelectron microscopy as described in (32). The anti-Tgl antibody was used at a concentration of 1:15000.

RESULTS

***An in-frame deletion mutant of* safA**. We started this work by constructing strains where we could analyze the dependency on Tgl for the extractability of SafA and C30. A previously described (32) *safA* mutant was generated by substituting the *safA* gene with that for spectinomycin resistance (Δ$safA_{spc}$). In the absence of SafA the assembly of the inner coat is defective. Consequently, Δ*safA* spores present a decrease in lysozyme resistance and the analysis of spore coat proteins resolved by SDS-PAGE shows that several proteins are less extractable or missing when compared to wild type spores (Figure 3.1A, compare first two lanes) (18,32). Complementation of the Δ$safA_{spc}$ mutant with a copy of *safA* by a single cross-over event at the *safA* locus restores spore resistance to lysozyme (32). However, the analysis of the pattern of spore coat proteins, as detected by SDS-PAGE, shows that the complemented strain does not recover the wild type phenotype. Rather, it resembles more the phenotype of the *safA* mutant (Figure 3.1A, first three lanes). We reasoned that the defects detected in the complemented strain were due to polar effects in downstream genes of the *safA* region. Thus, we decided to construct an in-frame deletion mutant of *safA*.

Residues of *safA* corresponding to codons 46 to 249 were deleted, eliminating production of SafA, C30 and N21 (29). The deletion of codons 46 to 249 also eliminates regions A and B of SafA (Figure 1.4B), which are important for its proper localization during sporulation, as they correspond to points of interaction with SpoVID (41). This in-frame *safA* deletion mutant displays the previously described characteristics of *safA* mutants (18,32), with the spores showing a decrease in lysozyme resistance (Figure 3.1B) and an altered pattern of extracted spore coat proteins as detected by SDS-PAGE (Figure 3.1A, Δ$safA_{46\text{-}249}$). Importantly, the introduction of the *safA*

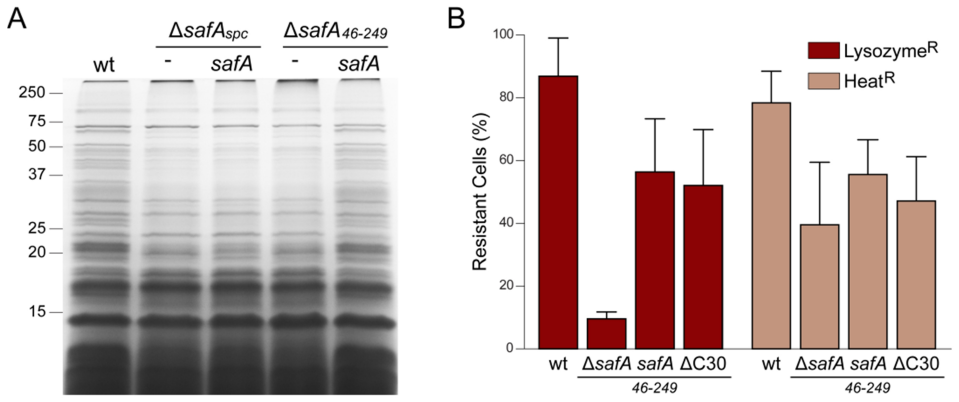


Figure 3.1. An in-frame *safA* deletion can be complemented to restore the wild type phenotype. Two *safA* mutant strains were constructed by a double crossover event. One led to the substitution of *safA* by the spectinomycin acetyltransferase gene (AOB68, (32)) (denoted as Δ*safA*$_{spc}$ in the figure), and the other led to an in-frame deletion of the sequence of *safA* that codifies for codons 46 to 249 (identified as Δ*safA*$_{46\text{-}249}$, corresponding to AH10297; note that this deletion abolishes the formation of SafA, C30 and N21). Strain Δ*safA*$_{spc}$ was complemented with *safA* by a single cross-over event at the *safA* locus (AH4745), while in strain Δ*safA*$_{46\text{-}249}$, *safA*, or derivatives, were inserted at the *amyE* locus (under the control of the native *safA* promoter). **A.** Analysis of coat proteins of purified spores. While both *safA* mutant strains present a similar phenotype, only the mutant strain with the in-frame deletion of *safA* can be complemented to restore the wild type (wt, MB24) pattern of spore coat proteins. **B.** The in-frame *safA* deletion mutant and its derivatives where tested for lysozyme and heat resistance. In accordance to previous studies (18,32), the absence of SafA impairs the spore resistance to lysozyme (one log decrease in cell survival). Note that complementation with *safA* (AH10302) or with *safA*$_{M161/164A}$ (represented as ΔC30; AH10383) restores the spore's resistance to lysozyme (in *safA*$_{M161/164A}$, the two mutations abolish the formation of the C30 form, but not the full length or N21 forms of SafA, hence referred has ΔC30).

wild type allele, at the *amyE* locus of the in-frame *safA* mutant, leads to spores with a wild type pattern of spore coat proteins (Figure 3.1A, last lane) and lysozyme resistance (Figure 3.1B). Because the Δ*safA*$_{46\text{-}249}$ s*afA* complementation strain behaved as the wild type in all the assays (Figure 3.1C), the in-frame *safA* deletion mutant background was adopted throughout this study.

Dependency of SafA and C30 on Tgl. We then tested the effect of the absence of Tgl on SafA and C30. As can be seen in Figure 3.2A, and in agreement with previous studies (27), in a *tgl* mutant background, the extractability of SafA from purified spores is greatly increased as detected by immunoblot analysis with anti-N21 antibody (which only detects SafA and N21) (compare lanes 1 and 2 and note that in Δ*tgl*, 1/10 of sample volume was applied to the gel). Also, in the absence of Tgl, the released

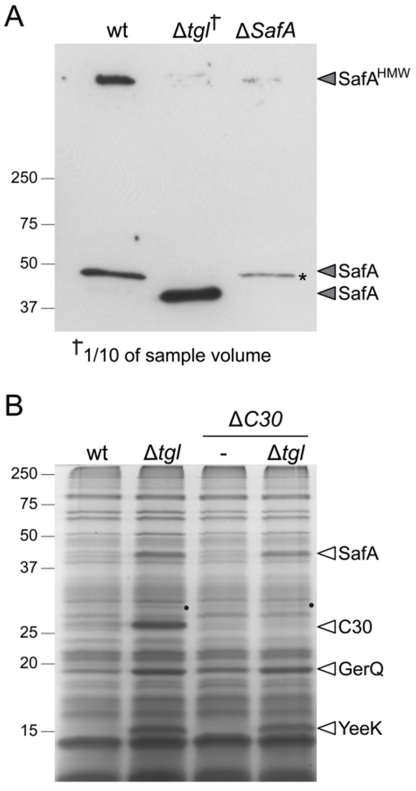


Figure 3.2. Confirmation of SafA and C30 as substrates of Tgl. Spore coat proteins were extracted from purified spores and resolved by SDS-PAGE. **A.** Spore coat proteins of wild type (wt, MB24), Δ*tgl* (AH2255) and Δ*safA* (AH10297) strains were analyzed by immunoblot with anti-N21 antibody (a polyclonal antibody raised against N21, which is able to recognize SafA and N21, but not C30). The site of migration of monomeric and high molecular weight products of SafA (SafA and SafA^{HMW}, respectively) is shown. It is very clear that in the absence of Tgl, SafA becomes much more extractable and it is detected as a faster migrating band, which most likely corresponds to the processed form of SafA (also note that for the Δ*tgl* sample, 1/10 of the volume of the other samples was used). The asterisk shows the site of migration of a cross-reactive unspecific species (that migrates very close to SafA, but not on the same site) **B.** The analysis of spore coat proteins shows that the absence of C30 (Δ*C30*, AH10383) does not affect the pattern of extracted spore coat proteins. The absence of C30 in a *tgl* mutant (Δ*C30* Δ*tgl*, AH10387) confirms that the more extractable ~25 kDa band detected in Δ*tgl* spores corresponds to C30. The black dots represent the presence of a more extract-able polypeptide in strains with a Δ*tgl* background; this protein presumably corresponds to unknown substrate of Tgl. Note that this band is very faint and not always easily detected. The site of migration of the molecular weight markers (in kDa) is represented on the left side of each panel.

SafA is present as a faster migrating species, possibly corresponding to its YabG-dependent processed form (as previously seen (27)). In the wild type, only the unprocessed form is detected.

Next, we turned to C30, which is produced by internal translation of *safA*, starting at Met164 (29). However, when Met164 is exchanged to Ala, C30 can be produced from Met161, while the exchange of both methionine residues to alanine abolishes the formation of C30 (29). Thus, the Δ*C30* strain was created by introducing *safA*$_{M161/164A}$ in the *amyE* locus of our *safA* in-frame deletion mutant. The absence of C30 does not alter the resistance properties of spores (Figure 3.1B and (29,51)) and does not lead to severe alterations in the pattern of spore coat proteins, as visualized by SDS-PAGE (Figure 3.2B). As such, we could confirm, by direct visualization of SDS-PAGE gels of spore coat proteins, that the more extractable ~25 kDa band present in *tgl* mutant spores corresponds to C30, as it is not present in the double

C30 tgl mutant (Figure 3.2B).

In Figure 3.2B, it is also possible to see an extra more extractable band in strains bearing a Δ*tgl* deletion mutation (black dots). This band is not always easily detected, but presumably corresponds to a fifth unidentified substrate of Tgl. We were, however, unable to identify this protein (see also Figure 3.4B). Finally, we also noticed that in the absence of C30, the extractability of GerQ appears slightly increased when compared to the wild type (compare first and third lanes in Figure 3.2B), and this possibly suggests that C30 is a cross-linking partner for GerQ.

Previous studies had shown that the extractability of SafA and C30 is increased in *tgl* mutant spores as detected by western blot analysis with an anti-SafA antibody (27). In all, and in extension of this previous report, our results strongly suggest that both C30 and SafA are direct substrates of Tgl.

The extractability of cortex-associated SafA and C30 is dependent on Tgl. As it had been previously suggested (27), our results indicate that both SafA and C30 are substrates of Tgl. However, previous studies have also indicated that SafA (and possibly C30) localizes to the cortex/inner coat interface (32). In addition, SafA has a LysM domain (peptidoglycan-binding domain (52)) in its N-terminal region. Thus, it is likely that SafA is associated, at least in part, with the cortex. We decided to test if in fact SafA is associated with the cortex compartment and if this association is dependent on Tgl.

Purified spores were subjected to a decoating regime which removes the extractable coat proteins (coat fraction, Figure 3.3) and exposes the cortex layer. The cortex is composed by peptidoglycan and incubation of decoated spores with lysozyme releases proteins presumably associated with the cortex layer (+lys, in Figure 3.3). Decoating of wild type spores removes essentially all of the extractable coat proteins and leaves behind a rind of insoluble coat material that remains following lysozyme digestion (53). As can be seen in Figure 3.3B, SafA and C30 were detected as species of about 45 kDa and 30 kDa in the coat fraction of wild type spores, as expected. However, both proteins were also detected in the lysozyme fraction (Figure 3.3B), which implies that SafA and C30 are associated with the cortex layer. Importantly, a bona fide coat protein, CotA (20), was only detected in the

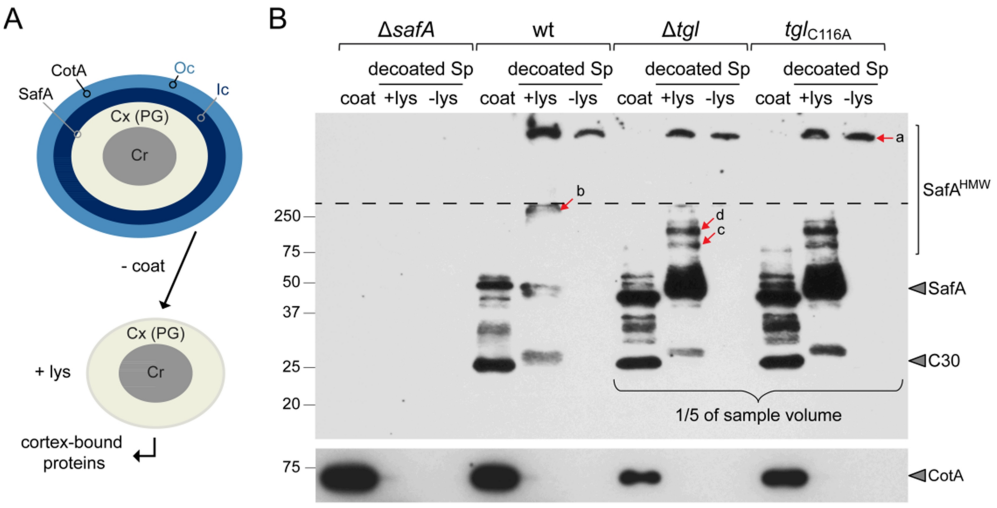


Figure 3.3. Effect of the absence of Tgl on the extractability of SafA and C30 in the different layers of the spore. Cultures of Δ*safA* (AH10297), wild type (wt, MB24), Δ*tgl* (AH2255), and Δ*tgl* tgl_{C116A} (AH10388) strains were grown in DSM, collected at hour 24 of sporulation, and the spores purified. **A and B.** Coat proteins of purified spores were extracted (referred as coat); the decoated spores have the cortex layer exposed which is composed of peptidoglycan (PG); the decoated spores were incubated at 37° C for 2 hr, with or without lysozyme (+lys, -lys, respectively in B), which releases cortex associated proteins. The localization of CotA and SafA are illustrated in A; CotA is a bonafide coat protein (20) present in the outer coat; SafA has been seen to localize at the cortex/inner coat interface (32). Cx, Ic and Oc refer to cortex, inner and outer coat layers, respectively. **B.** Samples were resolved by SDS-PAGE followed by immunoblot analysis with an anti-SafA or anti-cotA antibody. The sites of migration of $SafA^{HMW}$ (high molecular weight species detected with anti-SafA), SafA, C30 and CotA are identified. The dashed line represents the interface between the stacking and resolving gels. The molecular weight markers' (in kDa) position is represented on the left side of the panel. Note that the C30 released in the coat and lysozyme fractions migrates with two different apparent sizes, while SafA released in the coat fraction shows an altered migration pattern in the wild type when compared to *tgl* mutant or tgl_{C116A} strains.

coat fraction, as anticipated (Fig. 3.3B). No protein is presently known that serves as a specific marker for the spore cortex layer.

The C30 protein in the lysozyme fraction migrates slightly slower than the form found in the coat fraction (Figure 3.3B). It is possible that the coat-associated C30 results from processing by the coat-associated YabG protease, whereas the cortex-associated form remains uncleaved (27). In addition, high molecular weight forms, i.e., with apparent sizes larger than that of the full-length SafA protein, were detected in the lysozyme fraction ($SafA^{HMW}$, in Figure 3.3B). These species, detected above the 250 kDa marker (identified as b), or at the origin of the gel well (identified as a), were not found in the coat fraction. The species migrating at ~250 kDa (band b),

as well as SafA, and C30 were not detected when the decoated spores were re-extracted, without prior treatment with lysozyme (Fig. 3.3B, -lys lane). Because species corresponding to band a, are detected in decoated spores, with or without lysozyme treatment, it is likely that they correspond to the highly insoluble rind of coat material that remains after lysozyme treatment (53). All the signals detected represent forms of SafA, as they were not detected in extracts from spores of the *safA* mutant (Fig. 3.3B).

We take these observations as suggesting that two populations of SafA are found in spores, consistent with the immunoelectron localization of SafA at the inner coat/cortex interface and with the presence of a LysM domain in the protein (32). One population is mostly associated with the spore coat; while the other, which is only released by lysozyme digestion, may be more strongly associated with the spore cortex. This second fraction of SafA includes high molecular weight forms that resist reducing agents and SDS during extraction and SDS-PAGE (band b, in Figure 3.3B). C30 lacks a LysM domain and is not recruited to the developing spore in the absence of the full-length protein with which it interacts (54). C30 may, thus, be recruited to the lysozyme fraction by SafA.

We then accessed the effect of the absence of Tgl in the extractability of the fraction of SafA and C30 that is associated with the cortex layer of the spore. The results indicate that in Δ*tgl* spores the extractability of both SafA and C30 is significantly increased, not only in the coat, but also in the lysozyme treated fractions, when compared to wild type spores (Figure 3.3B, note that in the *tgl* mutant, 1/5 of the sample volume was used, when compared to the wild type or Δ*safA* strains). However, the Δ*tgl* mutation had a greater impact on the representation of SafA in the lysozyme fraction, as compared to the coat fraction (unlike C30, Figure 3.3B). Interestingly, the form of SafA around 250 kDa (band b) released by lysozyme treatment of wild type spores, was not detected in the same fraction of Δ*tgl* spores (Figure 3.3B). Instead, two new forms of SafA of about 75 and 100 kDa (identified as c and d) were detected in the Δ*tgl* mutant (Figure 3.3B). This suggests the involvement of Tgl in the cross-linking of SafA in the lysozyme fraction.

To test whether the increased extractability of SafA and C30 was due to the absence of Tgl activity and not an indirect effect of the absence of the polypeptide, we transferred to the non-essential *thrC* locus of the Δ*tgl* strain, an allele coding for an inactive form of the enzyme, tgl_{C116A} (Chapter2 and (26,55)). In agreement with previous results (26), Tgl^{C116A} was detected at similar levels as the wild type enzyme in purified spores (not shown). However, as for the Δ*tgl* mutant (above), the representation of SafA and C30 in the coat and lysozyme fractions also dramatically increased. For SafA, this effect was more pronounced in the lysozyme fraction, similar to what is detected in *tgl* mutant spores (Figure 3.3B). Except for high molecular weight material found at the origin of the gel, no other forms of SafA were found when the sediment of decoated spores was re-extracted, without prior treatment with lysozyme (Figure 3.3B).

As shown in Figures 3.2A and 3.3B, the SafA released in the coat fraction of purified spores of a *tgl* mutant (or tgl_{C116A}) migrates faster than the one released from wild type spores. This form most likely corresponds to a YabG-dependent processed form of SafA (27). However, cortex-associated SafA (+lys lanes in Figure 3.3B) of wild type, *tgl* mutant and tgl_{C116A} spores shows the same apparent size. It is, thus, possible that the differences in size detected for SafA, in the coat and lysozyme fractions of *tgl* mutant spores, is not due to processing versus unprocessing. Rather, the released cortex-associated SafA may correspond to processed SafA with an altered form or bound to small polypeptides or fragments of peptidoglycan. This would mean that the SafA which is easily extracted from purified spores (coat fraction), corresponds to the little unprocessed form of the protein that still remains in spores. This also raises the possibility that the coat and cortex-associated C30 species, which are detected as different migrating forms, also do not correspond to different processed forms of C30.

Overall, our results indicate that in the absence of active Tgl, SafA and C30 are more extractable from the coat and lysozyme fractions of spores. Because the levels of both proteins extracted from *tgl* mutant or tgl_{C116A} spores is far greater than what is detected for the wild type (in Figure 3.3B, for *tgl* and tgl_{C116A}, 1/5 of sample volume was applied to the gel, when compared to the wild type), it is reasonable to assume that Tgl is

responsible for the cross-linking of SafA and C30 into forms that resist extraction. Thus, the two proteins may be part of the insoluble coat material that is resistant to solubilization by different methods (22,53,56). This implies that Tgl is necessary for the enduring association of SafA and C30 within both coat and cortex layers, which is consistent with SafA localization at the inner coat/cortex interface (32). In any event, proper assembly of SafA and C30 requires the activity of Tgl.

Tgl is also associated with the cortex layer of spores. If Tgl is responsible for alterations in cortex-associated SafA and C30, one possibility is that Tgl itself may also be linked (at least in part) with the cortex. To test this, we analyzed the extractability of the enzyme in the different layers of the spore. We detected Tgl in the coat and lysozyme treated fractions of purified spores and to a less extent in the fraction of decoated spores not treated with lysozyme (Figure 3.4A). These results suggest that Tgl associates with both coat and cortex layers. Like for SafA (Figure 3.3B), we also detected high molecular weight products of Tgl (at the origin of the gel) but only in the fraction of decoated spores treated with lysozyme (Figure 3.4A, Tgl^{HMW}). This may indicate that Tgl itself forms multimeric forms and may be part of the highly insoluble coat material. Additionally, it has been seen that, *in vitro*, Tgl is able to cross-link itself (38), supporting this hypothesis. Lastly, we examined the extractability of Tgl from the different fractions of *safA* mutant spores. Strikingly, Tgl could only be detected in the coat fraction of *safA* mutant spores (Figure 3.4A). Moreover, the levels of Tgl in the coat fraction were only slightly reduced in comparison to wild type spores (Figure 3.4A). Together, the results suggest that Tgl is associated with both the coat and cortex layers of the spore. However, the association of Tgl with the cortex is reliant on *safA* expression.

We have also analyzed the coat and lysozyme extracted proteins by Coomassie staining after SDS-PAGE resolution. The analysis of the coat fraction shows an extra more extractable band that does not correspond to the known substrates of Tgl (black dot in the coat fraction of Figure 3.4B; see also Figure 3.2B), which may correspond to an additional unknown substrate of Tgl. The incubation of decoated spores with lysozyme releases

several proteins (Figure 3.4C). A comparison between wild type and *tgl* mutant spores shows that in the absence of Tgl two proteins seem more extractable (identified as black dots in Figure 3.4C), and while the protein migrating between 37 and 50 kDa may correspond to SafA, the second more extractable band does not appear to correspond to any of the known substrates of Tgl. This may indicate that Tgl is responsible for the cross-linking of unidentified proteins within the cortex layer. In the lysozyme fraction of the wild type strain, there is a more extractable protein whose identity is also unknown. In all, the results indicate that Tgl is most likely present in the coat and cortex layers where it is responsible for the cross-linking of several proteins.

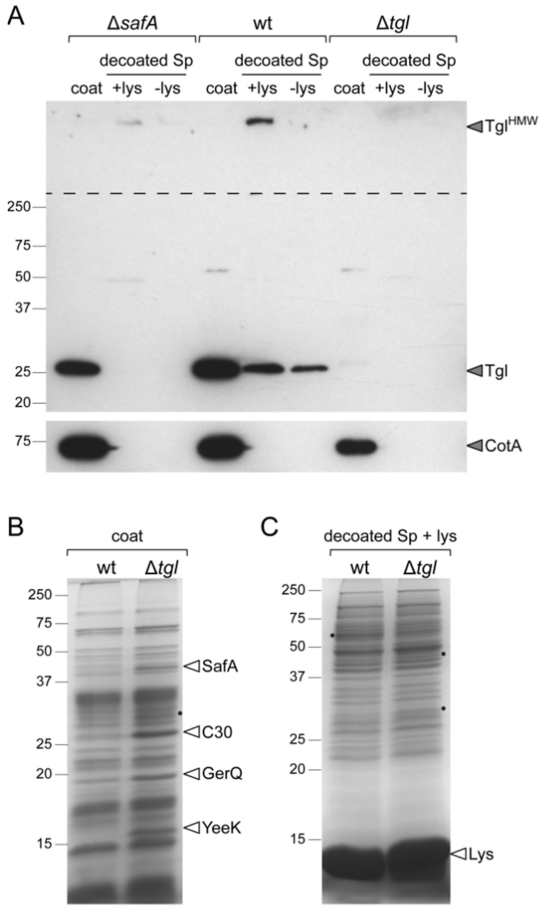


Figure 3.4. Tgl can be extracted from different layers of the spore. Cultures of Δ*safA* (AH10297), wild type (wt,MB24), and Δ*tgl* (AH2255) strains were grown in DSM for 24 hr and the spores purified. Spore coat proteins were extracted (coat) and the decoated spores were incubated at 37° C, in Tris buffer for 2 hr, with or without, lysozyme (lys). **A.** Samples were resolved by SDS-PAGE and analyzed by western-blot with anti-Tgl and anti-CotA antibodies. The site of migration of Tgl and CotA is represented. The dashed line indicates the interface between the stacking and resolving gels. **B and C.** Samples corresponding to the extracted coat proteins (**B**) and decoated spores treated with lysozyme (**C**) were resolved by SDS-PAGE and stained with Coomassie brilliant blue. The site of migration of the substrates of Tgl (in B) and of lysozyme (Lys, in C) is indicated by white arrowheads. The black dots indicate the presence of more extractable species, either in the wild type or Δ*tgl* strains, corresponding to proteins whose extractability is influenced by Tgl; note that some of these species may correspond to additional unidentified substrates of Tgl. The site of migration of the molecular weight markers (in kDa) is indicated on the left side of each panel.

Tgl localizes to the cortex, inner and outer coat layers. While Tgl appears to be associated with the coat and cortex layers it is still unknown whether Tgl localizes to both inner and outer coat layers. In order to answer this question we decided to analyze the localization of Tgl by immunoelectron microscopy with an anti-Tgl antibody. Spores of cultures grown in DSM for 24 hr after the initiation of sporulation were examined and, as can be seen in Figure 3.5, Tgl predominately localizes to the cortex or at the interface between the cortex and inner coat layers (yellow arrows in Figure 3.5), but it is also detected in the outer coat (green arrows in Figure 3.5). The predominant localization of Tgl in the cortex/inner coat interface is reminiscent of what has been seen for SafA (32). The minor localization of Tgl in the outer coat layer appears to be in contrast with previous studies that have identified CotE, the morphogenetic protein responsible for outer coat assembly (19,21), as a key determinant in the assembly of Tgl during sporulation (20,26,27). As such, we decided to investigate the assembly of Tgl during sporulation and its dependency on SafA and CotE.

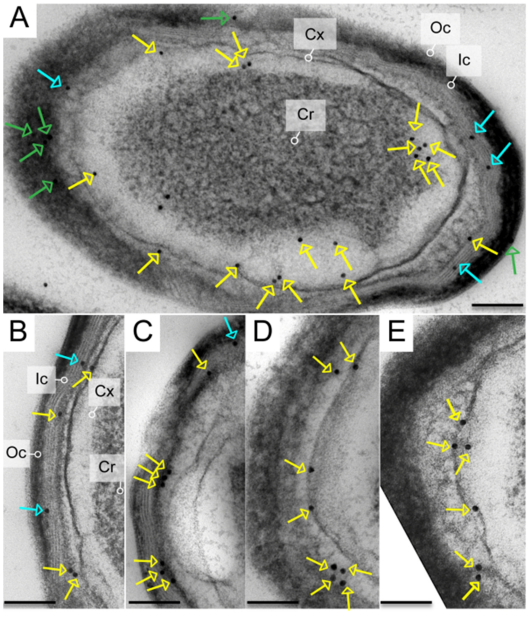


Figure 3.5. Tgl localizes mainly in the cortex and cortex/inner coat interface as detected by immunogold electron microscopy. Wild type, MB24, cultures were grown in DSM and collected 24 hr after the initiation of sporulation. Samples were analyzed by immunogold electron microscopy with an anti-Tgl antibody, and the presence of Tgl is indicated by the detection of the electron dense gold particles. The spore cortex (Cx), inner (Ic) and outer (Oc) coat layers are indicated. Gold-particles located in the cortex or cortex/inner coat interface are signaled by yellow arrows, while gold-particles detected in the inner and outer coat layers are pointed out by blue and green arrows, respectively. Scale bar corresponds to 200 nm.

***Construction of a functional tgl-cfp fusion protein*.** To examine the assembly of Tgl during sporulation by fluorescence microscopy, we sought to construct a functional Tgl-CFP fusion. Different *tgl-cfp* constructs were

made and the functionality of the fusion proteins was assessed by the examination of the profile of proteins extracted from spores, in comparison with the wild type and spores of a *tgl* deletion mutant. We introduced three different linkers between Tgl and CFP: 3N, a small linker composed of 3 asparagine residues; FL3, consisting of three repeats of the GGGGS sequence, which assumes a flexible conformation (57,58) and; HL4, comprised by four repeats of the EAAAK sequence, which forms a rigid α-helix (57,58) (for the complete sequence of the FL3 and HL4 linkers see materials and methods). All three *tgl-cfp* constructs, expressed from the *tgl* promoter, were inserted at the *amyE* locus of a *tgl* deletion mutant (strains AH10170, AH10271, and AH10272). In all cases, the resulting Tgl-CFP proteins appeared to be only partially functional, as C30 and GerQ were more extractable from spores of the strain expressing the fusion protein than in wild type spores (not shown for *tgl-3N-cfp* and *tgl-fl3-cfp*; for *tgl-hl4-cfp* see Figure 3.6A, fourth lane). However, while Tgl-3N-CFP and Tgl-Fl3-CFP appeared to be strongly processed, with the Tgl moiety detected to high levels by immunoblot analysis with an anti-Tgl antibody (not shown), the same was not true for Tgl-HL4-CFP (Figure 3.6B). Additionally, immunoblot analysis with an anti-Tgl antibody revealed that the levels of Tgl-HL4-CFP in purified spores were significantly lower than the levels of Tgl in wild type spores.

Expression of *tgl-hl4-cfp* from the *thrC* or *tgl* locus resulted in the same defects detected when *tgl-hl4-cfp* was expressed from the *amyE* locus; however introduction of *tgl* at the same loci fully restored the wild type phenotype from a *tgl* deletion mutant (Figure 3.6A-D). (NB: in all *tgl* complementation experiments referred to from now the *thrC* locus was used). Together, these observations raised the possibility that the Tgl-HL4-CFP protein was functional, but accumulated to reduced levels and/or showed impaired recruitment to the spore surface.

To investigate this point, we analyzed the levels of Tgl-HL4-CFP in the mother cell and forespore compartments during sporulation, by immunoblot analysis with an anti-Tgl antibody. In agreement with previous results (26), at hour 8 of sporulation, similar levels of Tgl were detected in the mother cell and forespore fractions of a wild type strain (Figure 3.6E). In

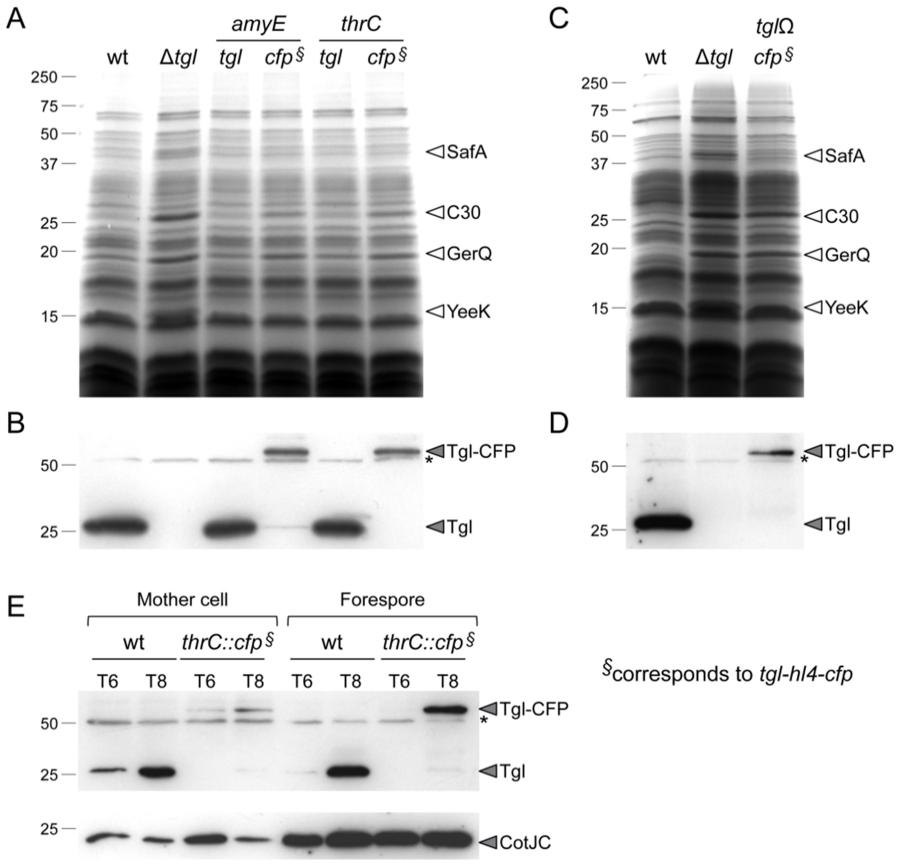


Figure 3.6. *tgl-hl4-cfp* fails to complement a Δ*tgl* mutation. *tgl* or *tgl-hl4-cfp* (*cfp*§) were introduced at the *amyE* (AH4599 and AH10271) or *thrC* locus (AH10307 and AH10308) of a *tgl* deletion mutant (AH2255), by marker replacement (panels A, B and E), or at the *tgl* locus, by Campbell recombination (AH10315) (C and D). **A-D.** Spores of the different strains were purified and the coat proteins extracted and resolved by SDS-PAGE followed by Coomassie staining (A and C) or immunoblot analysis with an anti-Tgl antibody (B and D). **E.** Samples of cultures of wild type and Δ*thrC*::*tgl-hl4-cfp* (AH10308) strains were collected 6 and 8 hr after the initiation of sporulation. The cells were lysed and separated into mother cell and forespore fractions and samples subject to immunoblot analysis with an anti-Tgl antibody. The membranes were then reprobed with an anti-CotJC antibody, as a loading control. White arrowheads indicate the position of the known substrates of Tgl. Gray arrowheads show the position of Tgl-HL4-CFP, Tgl or CotJC, while the asterisk shows the position of a non-specific cross-reactive species. Molecular weight markers, in kDa, are represented on the left side of each panel.

contrast, Tgl-HL4-CFP was mostly detected in the forespore fraction (Figure 3.6E). This suggested that Tgl-Hl4-CFP is produced at reduced levels and/or is unstable in the mother cell. Further, it suggested that increasing the production of Tgl-HL4-CFP could fully complement a *tgl* deletion mutation.

tgl transcription is under the control of σ^{K} (26,27,59-61). Positions 2 and

3 of the -35 region of σ^K-dependent promoters are usually occupied by cytosine/adenine and adenine, respectively (59). However, in the promoter of *tgl*, they are occupied by thymine (Figure 3.7A). In contrast, the -10 region of the *tgl* promoter is identical to the consensus sequence. In an attempt to increase transcription from the *tgl* promoter, positions 2 and 3, in the -35 element of the *tgl* promoter, were altered to cytosine and adenine, respectively (Figure 3.7A).

To test the effect of the alterations made to the *tgl* promoter we monitored the levels of Tgl-HL4-CFP, during sporulation and in purified spores, with an anti-Tgl or anti-GFP antibody (Figure 3.7). The results showed that the expression of *tgl-hl4-cfp* from the promoter with the consensus -35 element, significantly increases the levels of the fusion protein during sporulation and in purified spores, when compared with the expression of *tgl-hl4-cfp* from the native *tgl* promoter (Figure 3.7B,C,F,G). Importantly, the Tgl or CFP moieties were detected in low levels when compared to the full-length protein, indicating low levels of processing of Tgl-HL4-CFP (Figure 3.7B,C,F,G). Similarly, the expression of *tgl* from a consensus -35 promoter also increased the levels of the enzyme during sporulation and in purified spores, and in this case, the enzyme was always detected at higher levels than when compared to the wild type, MB24, strain (Figure 3.7D,F and repetitions of the same experiment).

The levels of Tgl-HL4-CFP in purified spores still appear considerably lower than those detected for the wild type strain or for the strain expressing *tgl* from the native promoter at the *thrC* locus (Figure 3.7F). However, the pattern of spore coat proteins extracted from spores of the strain expressing *tgl-hl4-cfp* from the consensual promoter is similar to that of the wild type (Figure 3.7E, compare lanes 1 and 6), indicating full restoration of the wild type phenotype. This in turn suggests that Tgl-HL4-CFP is largely functional. Interestingly, the expression of *tgl* from a consensus -35 region leads to an excess of enzyme in the spores when compared to the wild type strain which is reflected in the decreased extractability of two of the enzyme's substrates, C30 and GerQ (Figure 3.7E, compare first and third lanes).

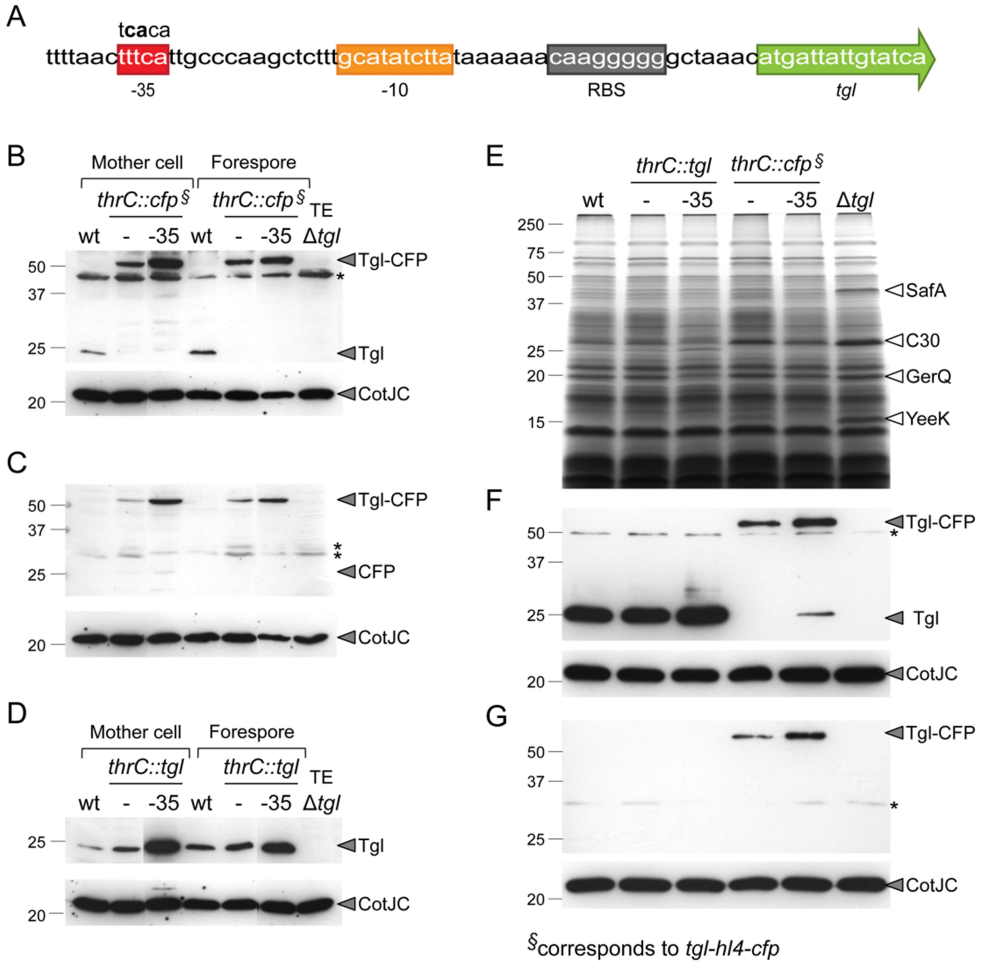


Figure 3.7. Production of Tgl-HL4-CFP from a consensus promoter fully complements a *tgl* deletion mutation. A. *tgl* promoter region with the -35, -10, RBS (ribosomal binding site) and coding regions identified (59). The nucleotides substituted for the construction of a strain with a -35 consensual element are shown in bold above the sequence. **B-G.** *tgl-hl4-cfp* (*cfp*§) or *tgl* were introduced at the *thrC* locus of a *tgl* mutant (AH2255), under the control of the native (AH10308 and AH10307) or consensual *tgl* promoter (AH10331 and AH10329) (represented as - or -35, respectively, while wt corresponds to the wild type strain, MB24). **B-D.** Samples were collected from cultures of the indicated strains, 7 hr after the initiation of sporulation in DSM. Whole cell total extracts (TE) were prepared and separated into mother cell and forespore fractions. Proteins in the whole cell extracts, mother cell, and forespore fractions were subject to immunoblot analysis with an anti-Tgl (**B** and **D**), or an anti-GFP antibody (**C**) and the membranes were reprobed with an anti-CotJC antibody, as a loading control. **E-G**. Coat proteins were extracted from spores purified from 24 hr DSM cultures. The coat proteins were resolved by SDS-PAGE and stained with Coomassie brilliant blue (**E**) or subject to immunoblot analysis with either an anti-Tgl (**F**), or an anti-GFP (**G**) antibody. For a loading control, the membranes were reprobed with an anti-CotJC antibody. White arrowheads in E, show the position of the known substrates of Tgl. Gray arrowheads show the position of Tgl-HL4-CFP, Tgl, CFP or CotJC. The asterisks mark the position of non-specific cross-reactive species. The position of the molecular weight markers (in kDa) is shown on the left side of panels B-G.

The expression of *tgl-hl4-cfp* from the consensus -35 promoter of *tgl* restores the wild type phenotype. This fusion, hereinafter termed Tgl-CFP for simplicity, was used to study the assembly of Tgl.

The assembly of Tgl during sporulation is dependent on SafA. We monitored the assembly pattern of our functional Tgl-CFP during sporulation in DSM. The expression of *tgl* is σ^K dependent and, as expected, we could only detect fluorescence in cells where engulfment had been completed (not shown). Furthermore, before phase dark forespores could be clearly visible we could detect a low fluorescence signal in the mother cell compartment but not in the forespore (not shown). Hence, we decided to only analyze cells with discernable phase dark or phase bright spores and focused on samples from 8 and 10 hours after the initiation of sporulation (Figure 3.8a,b). At these stages in sporulation, for about 98% of the cells analyzed, Tgl-CFP localizes at the forespore surface either as two caps of fluorescence (~60%) or as a ring of fluorescence around the forespore (~40%) (classes I and II in Figure 3.9A,B). A similar pattern is also detected in free spores 24 hours after the initiation of sporulation (Figures 3.8c and 3.9C).

To determine the dependency of Tgl-CFP on SafA for localization, *tgl-cfp* was introduced in a *safA* mutant strain. In the absence of SafA, about 40% of the population displayed a pattern of localization of Tgl-CFP similar to the wild type (Figure 3.8f,h, and classes I and II in Figure 3.9A,B). However, in the majority of the cells analyzed, Tgl-CFP tended to accumulate in the mother cell proximal pole of the forespore (Figure 3.8d,e,g, and class III in Figure 3.9A,B) or was dispersed in the mother cell cytoplasm, either homogeneously or in patches (Figure 3.8i and 3.9A,B, class IV). In all, classes III and IV represented 51% of the cells at both hours 8 and 10, as compared to 2% and 1% for the wild type. At T10, some cells displayed an accumulation of fluorescence in the mother cell distal pole of the forespore (Figure 3.8j; class V in Figure 3.9A,B), a pattern that was not detected at T8 or in the other strains analyzed. In all, the pattern of fluorescence detected in a *safA* mutant indicates that the assembly of Tgl-CFP is compromised.

Furthermore, whatever the pattern of Tgl-CFP detected in the *safA*

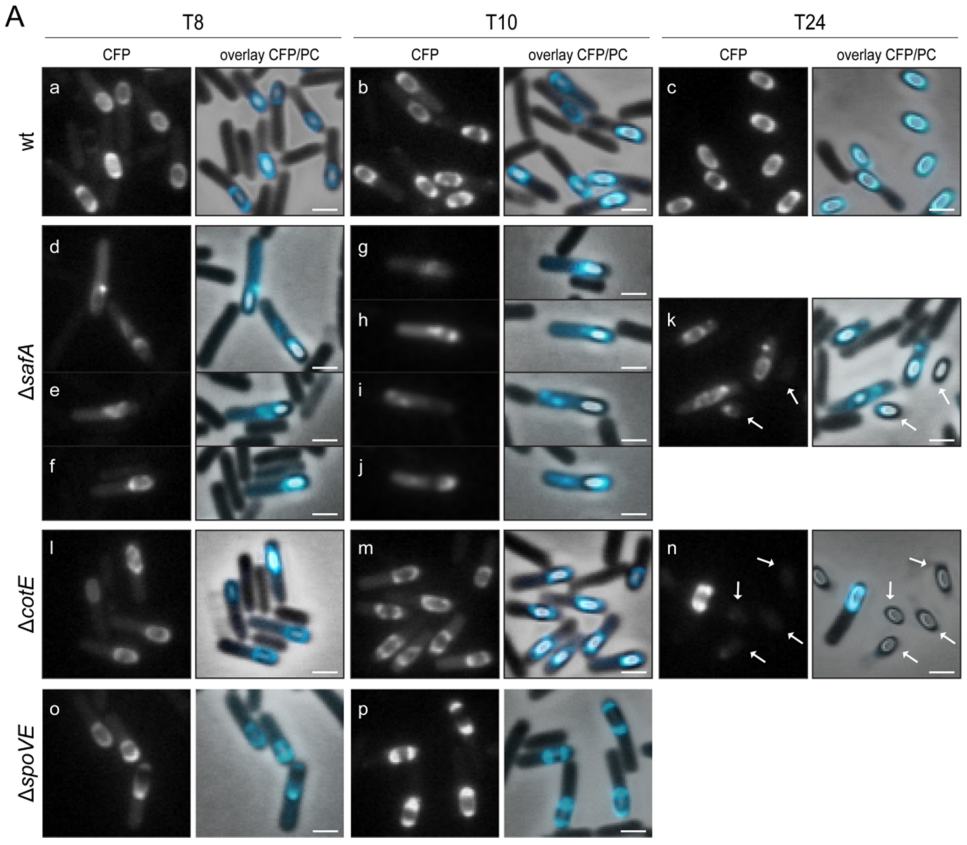
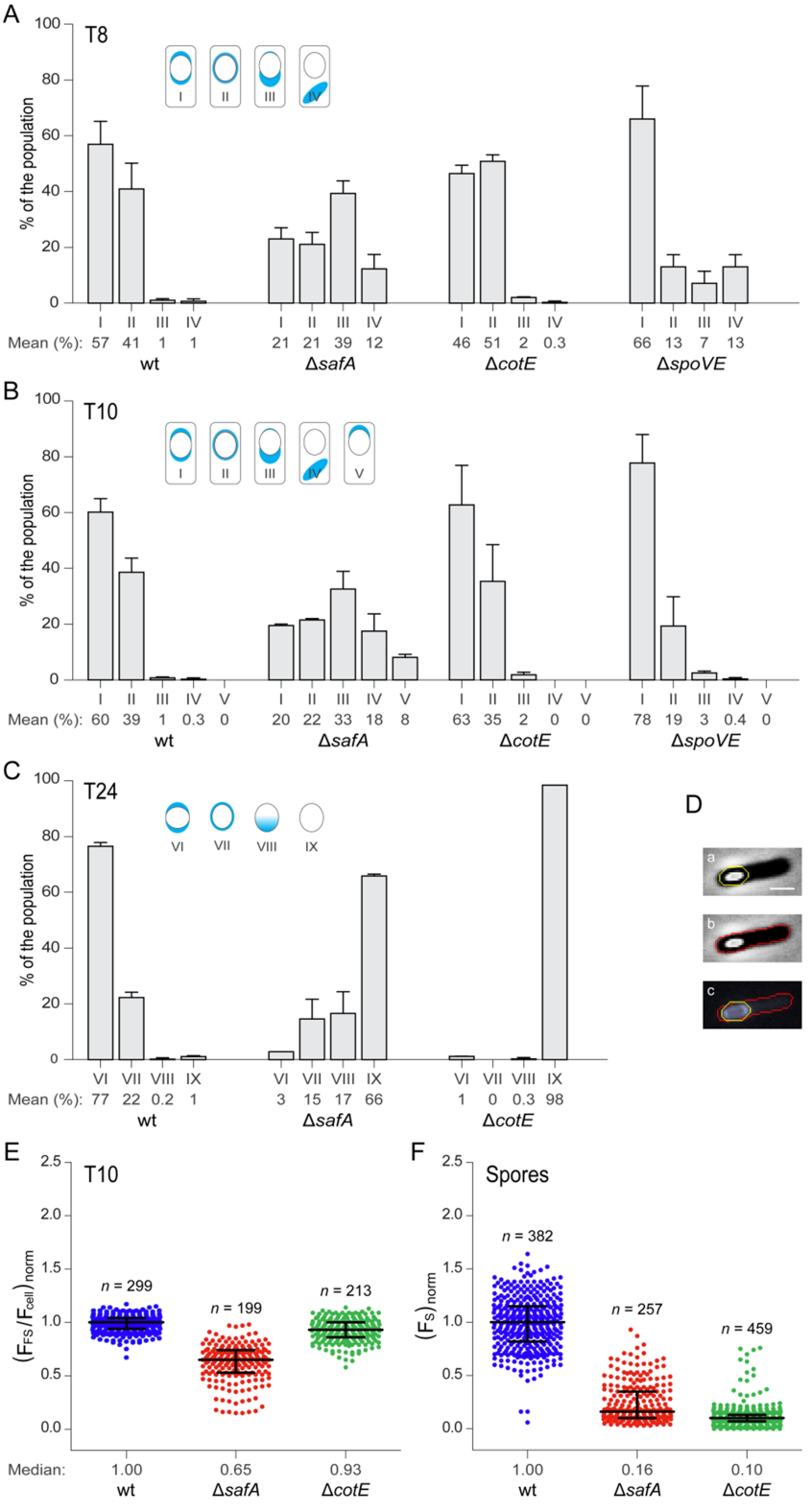


Figure 3.8 (*above*). The proper assembly of Tgl is dependent on SafA, CotE and cortex formation. Cells expressing a functional Tgl-CFP fusion protein in either a Δ*tgl* (AH10331), Δ*tgl* Δ*safA* (AH10357), Δ*tgl* Δ*cotE* (AH10341) or Δ*tgl* Δ*spoVE* (AH10336) background were grown in DSM and visualized by fluorescence microscopy. Samples were taken 8, 10 and 24 hr after the initiation of sporulation. Fields representative of the different patterns of fluorescence observed for each strain are shown. The CFP channel, and the overlay between the CFP and phase contrast (PC) channels are displayed. Scale bar, in white, corresponds to 1 µm. For the quantification analysis, please refer to Figure 3.9.

Figure 3.9. (*on the right*). The proper assembly of Tgl is dependent on SafA, CotE and cortex formation (*cont.*) A-C. Cells were analyzed and scored to different classes according to the pattern of fluorescence displayed (a diagram of the different patterns observed is represented); for T8 and T10, only cells with visible phase dark or phase bright spores were analyzed, while at T24 only free spores were examined. The mean and standard deviation for each class is represented. For all strains more than 150 cells were analyzed for each time point (see materials and methods). **D.** To quantify the signal of fluorescence, regions corresponding to the forespore (a) and whole cell (b) at T10, or free spores (similar to a) at T24, were determined using the images obtained in the phase contrast channel; these regions were then transferred to the CFP channel (c) and the intensity of fluorescence measured. **E and F.** The individual values of forespore/cell fluorescence ratio, or spore fluorescence, were determined and normalized by the median value of the wild type strain grown on the same day and in the same conditions (for details, see materials and methods). The normalized values were then plotted and the median and interquartile range is represented. Note that in E only cells containing phase bright forespores were analyzed. *n* corresponds to the number of cells/spores analyzed. The wild type values represented are from 3 independent experiments while the values for the other strains were obtained from 2 different independent experiments.


A
T8
% of the population
Mean (%): 57 41 1 1 | 21 21 39 12 | 46 51 2 0.3 | 66 13 7 13
wt
ΔsafA
ΔcotE
ΔspoVE
B
T10
Mean (%): 60 39 1 0.3 0 | 20 22 33 18 8 | 63 35 2 0 0 | 78 19 3 0.4 0
C
T24
Mean (%): 77 22 0.2 1 | 3 15 17 66 | 1 0 0.3 98
D
E
T10
(FFS/Fcell)norm
n = 299
n = 199
n = 213
Median: 1.00 0.65 0.93
F
Spores
(FS)norm
n = 382
n = 257
n = 459
1.00 0.16 0.10

mutant, the levels of fluorescence in the mother cell cytoplasm appeared always higher than in the wild type strain (compare a and b with d-j, in Figure 3.8). To test whether in the Δ*safA* strain Tgl-CFP was accumulating to higher levels in the mother cell compartment, we quantified the fluorescence signal in the forespore and whole cell (Figure 3.9D, see materials and methods for details). Due to methodology limitations only cells with phase bright forespores were analyzed. We determined the forespore/cell fluorescence ratio of individual cells and the values were normalized by the median value of the wild type strain that had been grown on the same day and in the same conditions as the strain being analyzed (the same was done to the individual wild type values themselves; note that the median and not the mean was used because the values did not show a normal distribution). The values obtained were plotted (Figure 3.9E) and it is clear that in the absence of SafA the forespore/cell fluorescence ratio is significantly lower than in the wild type (0.65 against 1.00, for the *safA* mutant and wild type strains, respectively). This implies that in the *safA* mutant, even in cells were Tgl-CFP localizes in a wild type-like pattern, part of the enzyme is retained in the mother cell compartment. Furthermore, note that the interquartile range in the *safA* mutant is wider than in the wild type, mirroring the widespread distribution of Tgl-CFP in different classes in the *safA* mutant, which does not happen in the wild type strain (Figure 3.9B).

The results indicate that SafA is important for the assembly of Tgl during sporulation and, in accordance, in the 24 hour sample, 77% of the free wild type spores, but only 3% of those of the *safA* mutant, showed a two caps pattern (Figure 3.9C, class VI). Rings of fluorescence around the spore where seen for 22% of the wild types spores and for 15% of the *safA* spores scored (Figure 3.9C, class VII). For the *safA* mutant, two additional classes were scored (Figure 3.9C): strong accumulation of fluorescence on one of the spore poles (class VIII, 17% of the spores) and no spore-associated fluorescence (class IX, 66% of the spores). In agreement with this analysis, the total fluorescence values detected for free spores, normalized to the median value of fluorescence detected in wild type spores, were considerably lower for *safA* spores (median value of 0.16, Figure 3.9F). That the most prevalent pattern seen for the wild type (two caps) was only residual for

safA spores and, that over 80% of the latter showed misassembled material or no fluorescence (classes VIII and IX), indicates that SafA is a key determinant for the localization of Tgl-CFP in mature spores. However, during sporulation, in a significant part of the population, Tgl-CFP is still able to assemble into the forespore according to a wild type-like pattern (~40%; classes I and II, in Figure 3.9A,B). Altogether, our results show that the assembly of Tgl-CFP is strongly, but not entirely, dependent on *safA.*

CotE is necessary for Tgl to remain associated with free spores. We next investigated whether the assembly of Tgl-CFP was dependent on *cotE*. Surprisingly, and in contrast to the results of previous studies (27,62), the pattern of Tgl-CFP localization during sporulation was not significantly affected by the *cotE* mutation at the time points examined (Figure 3.8l,m). Most of the cells at hour 8 showed the two cap or ring pattern (97% as compared to 98% for the wild type, sum of classes I and II in Figure 3.9A), a number that was maintained at hour 10 (99% as compared to 97% for the wild type; sum of classes I and II, Figure 3.9B). Importantly, accumulation of fluorescence at the spore poles or throughout the mother cell (classes III and IV seen for the *safA* mutant), was scarcely detected for the *cotE* mutant (Figure 3.9A,B). In agreement with these observations, in the absence of CotE only a slight decrease in the forespore/cell ratio of fluorescence was detected (0.93), as compared to the wild type (Figure 3.9E). Thus, the absence of CotE does not appear to have a great impact on the recruitment of Tgl to the surface of the developing spore.

In sharp contrast, the analysis of free spores in the 24 hr sample, reveals that 98% of the *cotE* spores showed no detectable fluorescence (Figure 3.8n; class IX in Figure 3.9C). The fluorescence signal was accordingly reduced as compared to wild type spores (Figure 3.9F, median value of 0.10 for *cotE* spores). These results establish that the recruitment of Tgl to the surface of the developing spore is largely independent on CotE. Rather, CotE is necessary for Tgl to remain associated with the released spores.

The localization of Tgl-CFP is affected by the inhibition of cortex synthesis. The partial requirement for *SafA*, for the localization of Tgl-CFP to the surface of the developing spore suggested that another factor controlled the

assembly of the enzyme. On the other hand, the localization of Tgl by immunogold labeling at the inner coat/cortex interface (above), together with the augmented extractability of Tgl by lysozyme digestion of decoated spores (Figures 3.4A and 3.5), converge to suggest an association of a fraction of Tgl with the cortex layer of spores. For this reason, we wanted to test whether synthesis of the cortex peptidoglycan was a requirement for the localization of Tgl-CFP during sporulation.

In *spoVE* mutants the coat is still assembled but cortex synthesis is blocked (15,63,64). We monitored the localization of Tgl-CFP, at hours 8 and 10 of sporulation, in a *spoVE* insertional mutant (Figure 3.8o,p). In 79% and 97% of the *spoVE* cells scored 8 and 10 hours after the onset of sporulation, respectively, Tgl-CFP formed two caps or a ring around the forespore, as compared with 98% (hour 8) and 99% (hour 10) for the wild type (Figure 3.9A,B, sum of classes I and II). In 20% (hour 8) and 3% (hour 10) of *spoVE* sporangia, Tgl-CFP accumulated in the mother cell proximal spore pole or dispersed throughout the mother cell cytoplasm (Figure 3.9A,B, sum of classes III and IV). As presented above, the sum of these two classes represented 51% of the sporangia at both hour 8 and 10 for the *safA* mutant. Thus, SafA makes a more important contribution than SpoVE to the targeting of Tgl-CFP to the developing spore. Nevertheless, SpoVE has a role, as classes III and IV are barely detected in the wild type. Additionally, while at T10, classes III and IV only represent 3% of the cells scored for the *spoVE* mutant, 78% of the cells are detected in class I and only 19% in class II (Figure 3.9B), while in the wild type these two classes correspond to 60 and 40%, respectively (Figure 3.9B). Thus, the absence of SpoVE appears to impair the localization of Tgl-CFP as a ring around the forespore.

In a *spoVE* mutant, the forespore and free spores never become phase bright due to the absence of the cortex. Thus, we could not quantify the forespore/cell ratio of fluorescence or the fluorescence detected in free spores, since our quantification method relies on the automatic detection of the retractile (fore)spore in phase contrast images. We also did not quantify the pattern of localization in free phase dark “sporelets” of the *spoVE* mutant as they are easily mistaken with cellular debris. Nonetheless, if Tgl-CFP localized in the released immature “sporelets” we should be able to

easily identify them in images taken with the CFP filter. Since this was not the case (not shown), it is likely that in free Δ*spoVE* "sporelets" Tgl is absent. Thus SpoVE is important for the long-term association of Tgl with free spores. The *spoVE*-specific effects on the localization of Tgl-CFP indicate that cortex formation contributes to the proper assembly of the fusion protein, even if only for a fraction of the transglutaminase, most likely the cortex-associated one.

DISCUSSION

We found that SafA, the morphogenetic protein responsible for inner coat assembly, is predominantly associated with the cortex layer in a form that resists extraction. The conversion of SafA into this form requires Tgl. Tgl itself, localizes to the coat in a manner that is partly independent on *safA* expression. However, Tgl also localizes to the cortex, and this localization requires *safA*. Therefore, the assembly of SafA and Tgl appears inter-dependent. We propose a model were Tgl is assembled in two steps (Figure 3.10): first the enzyme localizes in the forespore in a process that is partially independent on SafA; at this point Tgl is most likely present in the inner coat; then, Tgl is recruited to the cortex layer by SafA. The Tgl-dependent cross-linking in the coat of SafA, C30, GerQ, YeeK, and other unidentified proteins (Figures 3.2B and 3.4B) and the cross-linking of SafA, C30 and other putative substrates, in the cortex (Figure 3.4C), follows mother cell lysis (Figure 3.10B).

Our model for the assembly of Tgl is based on several observations. In wild type spores both SafA (and its shorter form C30), and Tgl are associated with the cortex and coat layers (Figures 3.3B, 3.4A and 3.5). In wild type spores, SafA/C30 are present in a form that is highly resistant to extraction (Figure 3.3B, band a) or as a high molecular weight species (band b in Figure 3.3B), whose formation requires the activity of Tgl, suggesting that SafA/C30 are cross-linked by Tgl. Furthermore, in *tgl* or tgl_{C116A} mutant spores, SafA is mostly extracted from the fraction of decoated spores treated with lysozyme, indicating that prior to cross-linking, SafA is mainly associated with the cortex; likewise by immunoelectron microscopy the majority of Tgl

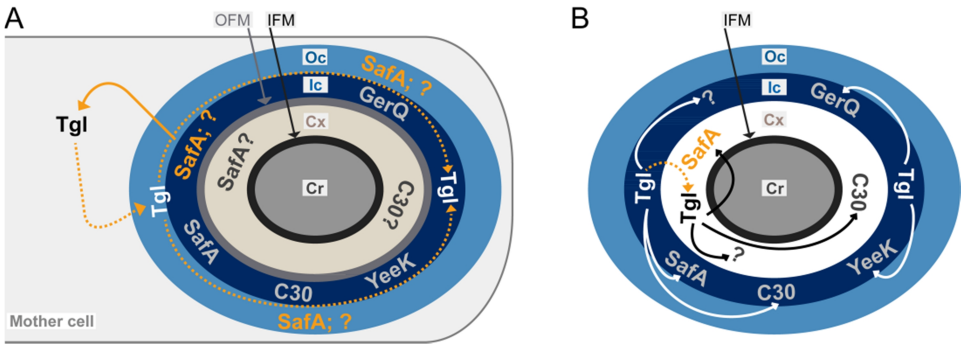


Figure 3.10. The assembly of Tgl is organized in two steps, and the final assembly state of the enzyme and SafA in the cortex, appears to be interdependent. A. During sporulation Tgl is synthesized in the mother cell compartment and recruited to the forespore in a partly SafA-independent mechanism. Tgl then localizes around the forespore and before cell lysis may be confined to the coat layers, while SafA and C30 may already localize to the cortex compartment. At this time, three of the substrates of Tgl, SafA, C30, and YeeK are processed in the coat layers in a YabG-dependent manner (not shown for simplicity). **B.** Following lysis of the mother cell, the free spore is released. It is possible that at this time the outer forespore membrane is absent and Tgl is recruited to the cortex by SafA. Alternatively, there is an unknown secretion system, which may function before mother cell lysis (not represented), which is responsible for the transport of Tgl and SafA into the cortex, across the outer forespores membrane. Cell lysis will induce the activity of Tgl (by an unknown mechanism) and the enzyme will cross-link SafA, C30, YeeK, and GerQ in the coat, and SafA and C30, in the cortex. The action of Tgl on the cortex localized SafA/C30 will transform the proteins into high molecular weight forms; however it is unknown whether Tgl cross-links SafA/C30 to other cortex-bound proteins and/or to the stem peptide of the cortex peptidoglycan. Our results also show that Tgl is likely responsible for the cross-linking of other unknown substrates in both coat and cortex and that C30 may be a cross-linking partner of GerQ. The solid arrows correspond to expected direct protein-protein interactions, while the dashed arrows represent localization dependencies. Cr, core; Cx, cortex; Ic, Inner coat; Oc, Outer coat; IFM, Inner forespore membrane; OFM, Outer forespore membrane.

is also detected in the cortex layer. On the other hand, when *safA* is eliminated, Tgl no longer associates with the cortex and is only detected in the coat (Figure 3.4A). This suggests that the subcellular localization of the enzyme within the cortex is mostly dependent on *safA* while its localization within the coat is not.

Tgl not only localizes to the cortex but also to the inner coat layer (usually at the inner coat/cortex interface) and to a minor extent in the outer coat layer of free spores (Figure 3.5). This widespread distribution of the enzyme within different layers of the spore suggests that the assembly of Tgl may be regulated by different factors. This is in line with the observation that in 40% of the population of a *safA* mutant, Tgl-CFP localizes in a wild-type pattern (Figure 3.9A,B). The additional factors that control

assembly of Tgl do not appear to be strongly dependent on proper cortex or outer coat formation, as in the absence of CotE or SpoVE, Tgl-CFP localization in sporulating cells is not strikingly impaired (this is especially true for cells analyzed at hour 10 of sporulation; Figures 3.8m,p and 3.9B). The previously seen requirement of CotE for the assembly of Tgl (26) or Tgl-CFP (20,27) is explained by the necessity of the outer coat for maintaining the association of Tgl with free spores (Figures 3.8n and 3.9C,F), as seen before for the inner coat protein OxdD (65). Likewise, SpoVE also appears to be important for the enduring association of Tgl with free spores.

One major question that arises from our model is how SafA and Tgl reach the cortex region. Engulfment completion creates a free protoplast surrounded by a double membrane. The intermembrane space determines the assembly place of the germ cell wall and the cortex layer. While the inner forespore membrane becomes the cytoplasmic membrane of the new cell generated by spore germination, the fate of the outer forespore membrane is not known, although some evidence suggests its absence from mature spores (66,67). SafA and Tgl could be transported across the outer forespore membrane to the cortex region, by an unknown mechanism. Recent work has, in fact, identified a novel sporulation-specific cell-cell secretion system. SpoIIQ and SpoIIIAH, synthesized in the forespore and mother cell compartments, respectively multimerize into a ring to form the core elements of a channel that spans across the inner and outer forespore membranes during engulfment (68-72). Although the SpoIIQ-SpoIIIAH channel is absent by the time Tgl is synthetized, a similar unknown mechanism might be involved in the transport of SafA and Tgl to the cortex region. Alternatively, the outer forespore membrane is absent by the time the proteins localize to the cortex, or the two pathways could be used.

In any event the activity of Tgl is detected late, following mother cell lysis (24,26). This implies that Tgl-dependent cross-linking in the spore is triggered by cell lysis by an unknown mechanism. It has also been proposed that Tgl may require processing of its substrates in a YabG-dependent manner. Nonetheless, while SafA, C30, and YeeK are processed during sporulation, their processed forms are detected about 6 hours after the initiation of sporulation ((34); not shown) soon after YabG synthesis (at T4,

(30)). Thus, substrate processing is not the only factor required for Tgl-dependent cross-linking.

The identity of the cross-linking partner(s) of SafA in the cortex is unknown. C30, whose extractability is also dependent on Tgl, is a candidate (Figure 3.3B). Another possibility is that Tgl cross-links SafA directly to the stem peptide of the cortex peptidoglycan. Tgl presents a unique fold different from all other transglutaminases and curiously it shows some similarity to peptidoglycan cysteine endopeptidases (Chapter 2). The cross-linking of proteins to peptidoglycan, catalyzed by a transglutaminase, has never been described; however most of the studies on transglutaminases concern the animal enzymes.

Our results show that C30 and SafA are associated with the cortex. Whether GerQ and YeeK are also associated with the cortex is at this time unknown. However, our results indicate that C30 may be a cross-linking partner of GerQ, as the absence of C30 increases the extractability of GerQ (Figure 3.2B) while the reverse is not true (26). Furthermore, initial attempts to produce a functional Tgl-CFP fusion protein also indicate that the extractability of C30 and GerQ from purified spores is linked (Figure 3.6A,C and data not shown). This dependency of GerQ on C30 suggests that GerQ may also be in the cortex, or at the cortex/inner coat interface as previously suggested (66). Furthermore, in *B. anthracis* Tgl appears to be responsible for the cross-linking of a cortex located protein (73). This indicates that Tgl may actually have a more global role in cementing the cortex/inner coat interface, which may account for the late activity of enzyme. Because some inner coat proteins are among the latest to be expressed during sporulation (9), the assembly of the inner coat must remain as a dynamic process until late in sporulation. Accordingly, Tgl should only act when “everything is in place” and as such the enzyme would be responsible for the stabilization of a pre-assembled complex (a process known as enzymatic spotwelding) (35).

The spotwelding role of Tgl appears to be analogous to that of Factor XIIIa (FXIIIa). FXIIIa is involved in the last stage of the blood coagulation cascade where it is mainly responsible for introducing cross-links between fibrin monomers in the blood clot. The assembly of the blood clot is

independent of FXIIIa activity but not its long lasting endurance and resistance to lytic enzymes (74). As such, similar to the role of Tgl, FXIIIa acts on a pre-assembled structure to confer increased resistance. Despite the fact that Tgl and FXIIIa appear to be structurally very different (Chapter 2), there are other similarities regarding their activity regulation. During blood clotting, the activation of FXIIIa is regulated by the assembly of its substrate, fibrin (35,74). Fibrin self-assembly on the other hand is activated by proteolysis (75). Thus, processing and self-assembly create the substrates for FXIIIa. Maturation of the capsid of certain virus is also noteworthy. Capsid maturation, often induced by proteolysis, leads to drastic conformational changes due to the reorganization of capsid proteins, which ultimately, increases the structural stability of the capsid (76,77). In the emblematic case of HK97, capsid stabilization is accompanied by the formation of intermolecular covalent isopeptide bonds between Lys and Asn residues of capsid proteins (78,79). Similar processes, processing of substrates and maturation of coat/cortex lattice, may concur to create the substrates for Tgl cross-linking.

Overall, our results suggest that the proper assembly of SafA and Tgl is interdependent. The final assembly state of both proteins appears to be regulated by the localization and activity of Tgl, which depends on SafA and can be described by a two-step model (Figure 3.10). Therefore, the assembly of SafA is auto-regulatory: SafA is necessary for the proper assembly and localization of Tgl, while the enzyme is necessary for the conversion of SafA into a form highly resistant to extraction. This auto-regulatory loop is likely important to maintain SafA as a dynamic recruiter in early stages of spore coat assembly and later (once the activity of Tgl is triggered) to be used as stabilizer by Tgl to increase the overall integrity of the spore coat.

REFERENCES

1. **Deneve C, Janoir C, Poilane I, Fantinato C, Collignon A.** 2009. New trends in *Clostridium difficile* virulence and pathogenesis. Int J Antimicrob Agents **33 Suppl 1:**S24-28.

2. **Weiner ZP, Glomski IJ.** 2012. Updating perspectives on the initiation of *Bacillus anthracis* growth and dissemination through its host. Infect Immun **80:**1626-1633.

3. **Henriques AO, Moran CP, Jr.** 2007. Structure, assembly, and function of the spore surface layers. Annual review of microbiology **61:**555-588.

4. **Kim H, Hahn M, Grabowski P, McPherson DC, Otte MM, Wang R, Ferguson CC, Eichenberger P, Driks A.** 2006. The *Bacillus subtilis* spore coat protein interaction network. Molecular microbiology **59:**487-502.

5. **Kuwana R, Kasahara Y, Fujibayashi M, Takamatsu H, Ogasawara N, Watabe K.** 2002. Proteomics characterization of novel spore proteins of *Bacillus subtilis*. Microbiology (Reading, England) **148:**3971-3982.

6. **Lai EM, Phadke ND, Kachman MT, Giorno R, Vazquez S, Vazquez JA, Maddock JR, Driks A.** 2003. Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*. Journal of bacteriology **185:**1443-1454.

7. **Henriques AO, Moran CP, Jr.** 2000. Structure and assembly of the bacterial endospore coat. Methods (San Diego, Calif **20:**95-110.

8. **Henriques AO, T. V. Costa, L. O. Martins, and R. Zilhão.** 2004. Functional architecture and assembly of the spore coat, p. 65-85. *In* Ezio Ricca AOH, and Simon M. Cutting (ed.), Bacterial Spores: Probiotics and Emerging Applications. Horizon Scientific Press, London, UK.

9. **McKenney PT, Driks A, Eichenberger P.** 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol **11:**33-44.

10. **Setlow P.** 2003. Spore germination. Current opinion in microbiology **6:**550-556.

11. **Higgins D, Dworkin J.** 2012. Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol Rev **36:**131-148.

12. **Piggot PJ, Hilbert DW.** 2004. Sporulation of *Bacillus subtilis*. Current opinion in microbiology **7:**579-586.

13. **Driks A.** 1999. *Bacillus subtilis* spore coat. Microbiol Mol Biol Rev **63:**1-20.

14. **Levin PA, Fan N, Ricca E, Driks A, Losick R, Cutting S.** 1993. An unusually small gene required for sporulation by *Bacillus subtilis*. Molecular microbiology **9:**761-771.

15. **Piggot PJ, Coote JG.** 1976. Genetic aspects of bacterial endospore formation. Bacteriological reviews **40:**908-962.

16. **Roels S, Driks A, Losick R.** 1992. Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. Journal of bacteriology **174:**575-585.

17. **Beall B, Driks A, Losick R, Moran CP, Jr.** 1993. Cloning and characterization of a gene required for assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **175:**1705-1716.

18. **Takamatsu H, Kodama T, Nakayama T, Watabe K.** 1999. Characterization of the *yrbA* gene of *Bacillus subtilis*, involved in resistance and germination of spores. Journal of bacteriology **181:**4986-4994.

19. **Driks A, Roels S, Beall B, Moran CP, Jr., Losick R.** 1994. Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. Genes & development **8:**234-244.

20. **McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z, Eichenberger P.** 2010. A distance-weighted interaction map reveals a previously uncharacterized layer of the *Bacillus subtilis* spore coat. Curr Biol **20:**934-938.

21. **Zheng LB, Donovan WP, Fitz-James PC, Losick R.** 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. Genes & development **2:**1047-1054.

22. **Pandey NK, Aronson AI.** 1979. Properties of the *Bacillus subtilis* spore coat. Journal of bacteriology **137:**1208-1218.

23. **Katsunori Kobayashi YK, Kiyoshi Miwa, Shigeru Yamanaka.** 1996. ε-(γ-Glutamyl)lysine cross-links of spore coat proteins and transglutaminase activity in *Bacillus subtilis*. FEMS microbiology letters **144:**157-160.

24. **Ragkousi K, Setlow P.** 2004. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. Journal of bacteriology **186:**5567-5575.

25. **Suzuki S, Izawa Y, Kobayashi K, Eto Y, Yamanaka S, Kubota K, Yokozeki K.** 2000. Purification and characterization of novel transglutaminase from *Bacillus subtilis* spores. Bioscience, biotechnology, and biochemistry **64:**2344-2351.

26. **Zilhao R, Isticato R, Martins LO, Steil L, Volker U, Ricca E, Moran CP, Jr., Henriques AO.** 2005. Assembly and function of a spore coat-associated transglutaminase of *Bacillus subtilis*. Journal of bacteriology **187:**7753-7764.

27. **Kuwana R, Okuda N, Takamatsu H, Watabe K.** 2006. Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. Journal of biochemistry **139:**887-901.

28. **Monroe A, Setlow P.** 2006. Localization of the transglutaminase cross-linking sites in the *Bacillus subtilis* spore coat protein GerQ. Journal of bacteriology **188:**7609-7616.

29. **Ozin AJ, Costa T, Henriques AO, Moran CP, Jr.** 2001. Alternative translation initiation produces a short form of a spore coat protein in *Bacillus subtilis*. Journal of bacteriology **183:**2032-2040.

30. **Takamatsu H, Kodama T, Imamura A, Asai K, Kobayashi K, Nakayama T, Ogasawara N, Watabe K.** 2000. The *Bacillus subtilis* yabG gene is transcribed by SigK RNA polymerase during sporulation, and *yabG* mutant spores have altered coat protein composition. Journal of bacteriology **182:**1883-1888.

31. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2010. Localization of proteins to different layers and regions of *Bacillus subtilis* spore coats. Journal of bacteriology **192:**518-524.

32. **Ozin AJ, Henriques AO, Yi H, Moran CP, Jr.** 2000. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **182:**1828-1833.

33. **Moir A.** 1981. Germination properties of a spore coat-defective mutant of *Bacillus subtilis*. Journal of bacteriology **146:**1106-1116.

34. **Takamatsu H, Imamura A, Kodama T, Asai K, Ogasawara N, Watabe K.** 2000. The yabG gene of *Bacillus subtilis* encodes a sporulation specific protease which is involved in the processing of several spore coat proteins. FEMS microbiology letters **192:**33-38.

35. **Lorand L, Graham RM.** 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature reviews **4:**140-156.

36. **Pasternack R, Dorsch S, Otterbach JT, Robenek IR, Wolf S, Fuchsbauer HL.** 1998. Bacterial pro-transglutaminase from *Streptoverticillium mobaraense*-purification, characterisation and sequence of the zymogen. European journal of biochemistry / FEBS **257:**570-576.

37. **Klöck C, Khosla C.** 2012. Regulation of the activities of the mammalian transglutaminase family of enzymes. Protein Science **21:**1781-1791.

38. **Placido D, Fernandes CG, Isidro A, Carrondo MA, Henriques AO, Archer M.** 2008. Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. Protein expression and purification **59:**1-8.

39. **Nicholson WL, Setlow P.** 1990. Sporulation, germination and outgrowth, p. 391-450. *In* Cutting CRHaSM (ed.), Molecular biology methods for *Bacillus*. John Wiley & Sons, Ltd., Chichester.

40. **Cutting SM, Vander Horn PB.** 1990. Genetic Analysis, p. 53-54. *In* Cutting CRHaSM (ed.), Molecular biology methods for *Bacillus*. John Wiley & Sons, Ltd., Chichester.

41. **Costa T, Isidro AL, Moran CP, Jr., Henriques AO.** 2006. Interaction between coat morphogenetic proteins SafA and SpoVID. Journal of bacteriology **188:**7731-7741.

42. **Karow ML, Piggot PJ.** 1995. Construction of *gusA* transcriptional fusion vectors for *Bacillus subtilis* and their utilization for studies of spore formation. Gene **163:**69-74.

43. **Doan T, Marquis KA, Rudner DZ.** 2005. Subcellular localization of a sporulation membrane protein is achieved through a network of interactions along and across the septum. Molecular microbiology **55:**1767-1781.

44. **Geiser M, Cebe R, Drewello D, Schmitz R.** 2001. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. BioTechniques **31:**88-90, 92.

45. **Guerout-Fleury AM, Frandsen N, Stragier P.** 1996. Plasmids for ectopic integration in *Bacillus subtilis*. Gene **180:**57-61.

46. **Zilhao R, Serrano M, Isticato R, Ricca E, Moran CP, Jr., Henriques AO.** 2004. Interactions among CotB, CotG, and CotH during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **186:**1110-1119.

47. **Henriques AO, Beall BW, Moran CP, Jr.** 1997. CotM of *Bacillus subtilis*, a member of the alpha-crystallin family of stress proteins, is induced during development and participates in spore outer coat formation. Journal of bacteriology **179:**1887-1897.

48. **Costa T, Serrano M, Steil L, Volker U, Moran CP, Jr., Henriques AO.** 2007. The timing of cotE expression affects *Bacillus subtilis* spore coat morphology but not lysozyme resistance. Journal of bacteriology **189:**2401-2410.

49. **Real G, Fay A, Eldar A, Pinto SM, Henriques AO, Dworkin J.** 2008. Determinants for the subcellular localization and function of a nonessential SEDS protein. Journal of bacteriology **190:**363-376.

50. **Seyler RW, Jr., Henriques AO, Ozin AJ, Moran CP, Jr.** 1997. Assembly and interactions of *cotJ*-encoded proteins, constituents of the inner layers of the *Bacillus subtilis* spore coat. Molecular microbiology **25:**955-966.

51. **Takamatsu H, Imamura D, Kuwana R, Watabe K.** 2009. Expression of yeeK during *Bacillus subtilis* sporulation and localization of YeeK to the inner spore coat using fluorescence microscopy. Journal of bacteriology **191:**1220-1229.

52. **Buist G, Steen A, Kok J, Kuipers OP.** 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. Molecular microbiology **68:**838-847.

53. **Ghosh S, Setlow B, Wahome PG, Cowan AE, Plomp M, Malkin AJ, Setlow P.** 2008. Characterization of spores of *Bacillus subtilis* that lack most coat layers. Journal of bacteriology **190:**6741-6748.

54. **Ozin AJ, Samford CS, Henriques AO, Moran CP, Jr.** 2001. SpoVID guides SafA to the spore coat in *Bacillus subtilis*. Journal of bacteriology **183:**3041-3049.

55. **Kobayashi K, Hashiguchi K, Yokozeki K, Yamanaka S.** 1998. Molecular cloning of the transglutaminase gene from *Bacillus subtilis* and its expression in *Escherichia coli*. Bioscience, biotechnology, and biochemistry **62:**1109-1114.

56. **Goldman RC, Tipper DJ.** 1978. *Bacillus subtilis* spore coats: complexity and purification of a unique polypeptide component. Journal of bacteriology **135:**1091-1106.

57. **Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T.** 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein engineering **14:**529-532.

58. **Arai R, Wriggers W, Nishikawa Y, Nagamune T, Fujisawa T.** 2004. Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. Proteins **57:**829-838.

59. **Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS biology **2:**e328.

60. **Steil L, Serrano M, Henriques AO, Volker U.** 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. Microbiology (Reading, England) **151:**399-420.

61. **Zheng L, Halberg R, Roels S, Ichikawa H, Kroos L, Losick R.** 1992. Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes. Journal of molecular biology **226:**1037-1050.

62. **McKenney PT, Eichenberger P.** 2012. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. Molecular microbiology **83:**245-260.

63. **Vasudevan P, Weaver A, Reichert ED, Linnstaedt SD, Popham DL.** 2007. Spore cortex formation in *Bacillus subtilis* is regulated by accumulation of peptidoglycan precursors under the control of sigma K. Molecular microbiology **65:**1582-1594.

64. **Coote JG.** 1972. Sporulation in *Bacillus subtilis*. Characterization of oligosporogenous mutants and comparison of their phenotypes with those of asporogenous mutants. J Gen Microbiol **71:**1-15.

65. **Costa T, Steil L, Martins LO, Volker U, Henriques AO.** 2004. Assembly of an oxalate decarboxylase produced under sigmaK control into the *Bacillus subtilis* spore coat. Journal of bacteriology **186:**1462-1474.

66. **Setlow P.** 2014. Germination of Spores of Bacillus Species: What We Know and Do Not Know. Journal of bacteriology **196:**1297-1305.

67. **Leggett MJ, McDonnell G, Denyer SP, Setlow P, Maillard JY.** 2012. Bacterial spore structures and their protective role in biocide resistance. Journal of applied microbiology **113:**485-498.

68. **Levdikov VM, Blagova EV, McFeat A, Fogg MJ, Wilson KS, Wilkinson AJ.** 2012. Structure of components of an intercellular channel complex in sporulating *Bacillus subtilis*. Proceedings of the National Academy of Sciences **109:**5441-5445.

69. **Meisner J, Maehigashi T, Andre I, Dunham CM, Moran CP, Jr.** 2012. Structure of the basal components of a bacterial transporter. Proceedings of the National Academy of Sciences of the United States of America **109:**5446-5451.

70. **Camp AH, Losick R.** 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*. Genes & development **23:**1014-1024.

71. **Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Jr., Rudner DZ.** 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. PLoS Genet **5:**e1000566.

72. **Meisner J, Wang X, Serrano M, Henriques AO, Moran CP, Jr.** 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. Proceedings of the National Academy of Sciences of the United States of America **105:**15100-15105.

73. **Moody KL, Driks A, Rother GL, Cote CK, Brueggemann EE, Hines HB, Friedlander AM, Bozue J.** 2009. Processing, assembly and localization of a *Bacillus anthracis* spore protein. Microbiology (Reading, England) **156:**174-183.

74. **Lorand L.** 2001. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. Ann N Y Acad Sci **936:**291-311.

75. **Yang Z, Mochalkin I, Doolittle RF.** 2000. A model of fibrin formation based on crystal structures of fibrinogen and fibrin fragments complexed with synthetic peptides. Proceedings of the National Academy of Sciences of the United States of America **97:**14156-14161.

76. **Domitrovic T, Movahed N, Bothner B, Matsui T, Wang Q, Doerschuk PC, Johnson JE.** 2013. Virus Assembly and Maturation: Auto-Regulation through Allosteric Molecular Switches. Journal of molecular biology **425:**1488-1496.

77. **Mateu MG.** 2013. Assembly, stability and dynamics of virus capsids. Archives of biochemistry and biophysics **531:**65-79.

78. **Duda RL.** 1998. Protein chainmail: catenated protein in viral capsids. Cell **94:**55-60.

79. **Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE.** 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science (New York, N.Y **289:**2129-2133.

Chapter 4

The assembly of a sporulation transglutaminase is driven by its substrates

This Chapter contains data to be published in: Catarina G. Fernandes, Ana L. Sousa, Carolina Freitas, Erin Tranfield and Adriano O. Henriques. 2014. *The assembly of a sporulation enzyme is driven by its substrates*. To be submitted to The Journal of Bacteriology.

The author of this dissertation performed all the experiments described in this chapter.

ABSTRACT

SafA, the morphogenetic protein responsible for the assembly of the inner coat of *Bacillus subtilis* spores, is an important but not unique determinant for the assembly of the Tgl transglutaminase. Tgl catalyzes the formation of ε-(γ-glutamyl)lysyl isopeptide bonds between protein bound glutaminyl and lysyl residues, resulting in protein cross-linking. At least four proteins, including SafA and its short form C30, produced by internal translation, as well as GerQ and YeeK, are more extractable from *tgl* mutant spores and therefore, are thought to be substrates of the enzyme. We show that the Tgl substrates have redundant roles in the recruitment of the enzyme to the spore coat. We also show that residues in Tgl, thought to be important for substrate recognition, are crucial for the assembly of the enzyme. We identified the glutamine and lysine substrate accepting sites (Q and K sides, respectively) at opposing ends of a tunnel that traverses the enzyme and that harbors the Tgl active site. We show that residues on the Q side of the enzyme are more important for the recruitment of Tgl than those on the K side. Because the Q donor substrate forms an acyl-intermediate with the enzyme prior to reaction with the K donor substrate, we propose that the first step in the reaction mechanism is also a key step in the recruitment of Tgl to the forespore. Consistent with the model that Tgl is recruited largely through enzyme-substrate interactions we show that a mutant form of the enzyme, Tgl^{W149A}, with altered activity against C30 *in vitro* is impaired in its assembly *in vivo*. Moreover, in a strain in which C30 is produced but not assembled, Tgl^{wt}, but not Tgl mutants with substitutions in Q side residues, co-localizes with C30 in the mother cell cytoplasm. The dependency of Tgl on the assembly of its substrates may ensure that the enzyme cross-links the correct proteins to stabilize a pre-formed scaffold.

INTRODUCTION

One of the fundamental questions in cell biology is how macromolecular structures are formed. One such example is the formation of the outermost layer of *Bacillus subtilis* endospores, denoted the coat. The coat is formed through the ordered assembly of more than 70 different proteins (1-4)

which in the end leads to the formation of a multiprotein structure that confers resistance to extremes of physical and chemical insults, and mediates the interactions of the spore with its immediate environment (1,5). How the assembly of the coat proteins is determined is still not completely understood and it's likely that different factors are important for different proteins. In this work we examined the assembly of Tgl, a spore coat enzyme involved in the cross-linking of coat proteins.

Species of the *Bacillus* and *Clostridium* genera, and related species, have the ability to switch from growth into a developmental process denominated sporulation, which will lead to the creation of a highly resistance dormant cell called spore. *B. subtilis'* spores are composed by three main concentric compartments: the core, which contains a copy of the genome; the cortex, composed by a modified form of peptidoglycan, which surrounds the core; and the coat, the outermost spore layer, that surrounds the cortex (6,7). The sporulation process starts with the asymmetric division of the cell into two different sized compartments denoted mother cell (larger) and forespore (smaller). The formation of the asymmetric septum triggers the expression of sporulation specific genes. The mother cell and the forespore will have different genetic programs which are controlled by the sequential activation of distinct sigma factors of RNA polymerase. Once the asymmetric septum is formed, σ^F is activated in the forespore followed by the activation of σ^E in the mother cell (8). The expression of genes coding for early spore coat proteins will be induced at this time under the control of σ^E (5). After asymmetric septum completion, the two membranes of the septum migrate around the forespore until it becomes a free protoplast inside the mother cell, in a process known as engulfment. Engulfment completion activates the transcription of a different set of genes under the control of σ^G and σ^K, in the forespore and mother cell, respectively (8,9). Late spore coat proteins are at this time transcribed in the mother cell compartment under the control of σ^K (5). Engulfment completion also triggers cortex formation and phase dark spores inside the mother cell become distinguishable by light microscopy. This is followed by cortex maturation processes leading to the appearance of phase bright forespores. Soon after, the sporulation process is finalized by mother cell lysis and spore

release.

The analysis of the coat structure by electron and fluorescence microscopy shows that it is composed by a basement layer which sits on the cortex and is surrounded by an inner and outer coat layers and finally by the crust (5). Although it is not completely known how the spore coat proteins are orderly assembled, a small subset of coat proteins, called morphogenetic proteins, have a severe impact on coat formation (1,5,10), and the formation of each of the layers of the coat is largely dependent on a specific morphogenetic protein (5). The assembly of the basement layer is determined by SpoIVA (11), SafA and CotE, are necessary for inner and outer coat formation, respectively (12-15), and CotX, CotY and CotZ (individually or combined), are responsible for the formation of the crust (14,16). While the above mentioned morphogenetic proteins create a scaffold for the construction of the different coat layers, two others, SpoVM and SpoVID, are determinant for the localization of spore coat proteins as a shell that surrounds the forespore in a process known as encasement (17). Although morphogenetic proteins drive spore coat assembly, isolated protein-protein interactions are also important for the assembly of distinct spore coat proteins. For example, CotJA and CotJC are thought to form a complex before assembly into the forespore which is essential for the localization of both proteins in the spore coat (18). It has also been seen that the assembly into the spore coat of CwlJ, a cortex-lytic enzyme (19), is dependent on GerQ, an inner coat protein (20).

In this work, we looked at the assembly dependencies of Tgl, a spore coat transglutaminase. Transglutaminases generate post-translational modifications in proteins by catalyzing the formation of covalent bonds between the side chains of glutaminyl and lysyl residues, leading to protein cross-linking (21). In previous work, we have seen that Tgl localizes mainly to the cortex and inner coat (Chapter 3). While the subcellular localization of Tgl in the cortex appears dependent on *safA*, the recruitment of the enzyme to the forespore during sporulation is only partly dependent on it (Chapter 3).

Using a functional Tgl-CFP fusion protein, we found that the substrates of Tgl are key factors in the recruitment of the enzyme to the surface of the

developing spore. We have also uncovered residues in the vicinity of the active site of Tgl that are important for the activity of the enzyme, most likely because they mediate enzyme-substrate interactions. We established that some of those residues are critical for the assembly of the enzyme during sporulation, indicating the involvement of enzyme-substrate-type interactions in the process. We suggest that the dependency of Tgl on its substrates is used to accurately direct the enzyme to the covalent fortification of a pre-assembled protein scaffold.

MATERIALS AND METHODS

Bacterial strains, media and general methods. *Escherichia coli* DH5α was used for plasmid constructions. The strains of *Bacillus subtilis* used in this study were derived from the wild type MB24 strain. Luria-Bertani medium was used for routine growth of *E. coli* and *B. subtilis* strains. For construction of plasmids, NZYDNAChange polymerase (NZYTech) was used for the amplification of fragments, while Phusion high-fidelity DNA polymerase (Finnzymes) was used when whole plasmids where amplified. Restriction enzymes from Fermentas were used. With the exception of plasmids created by sub-cloning strategies, all plasmids constructed were sequenced. For the construction of *B. subtilis* strains, insertions of genes at the *amyE* locus were confirmed by loss of extracellular amidase activity; insertions at other genes loci were confirmed by PCR (except when otherwise stated or when a resistance cassette gene was substituted). All strains, plasmids and primers used in this study are listed in Tables 4.1, 4.2, and 4.3, respectively.

Construction of the* yeeK *mutant. The *yeeK* gene, including upstream and downstream regions, was amplified with primers yeeK-409D and yeeK+828R using DNA from MB24 as template. The PCR product was digested with *Hind*III and cloned in the *Hind*III and *Sma*I sites of pUC18 (New England Biolabs), resulting in plasmid pCF62. Then, the chloramphenicol resistance cassette was amplified from pMS38 (22) with primers yeeK+260 cat-296 D and yeeK+170 cat+125 R. This PCR fragment was used as a megaprimer (23) to amplify pCF62, creating pCF65 which contains the coding region of *yeeK* interrupted by the *cm* gene and flanked by the upstream and downstream

regions of the *yeeK* gene. The introduction of the *cm* gene also leads to the elimination of 68 bp of the coding region of *yeeK* (base pairs 370 to 438). Plasmid pCF65 was used to transform MB24, by a double cross-over event, to create the *yeeK* mutant strain, AH10281. pCF62 and pCF65 presented strange forms as visualized by ethidium bromide agarose gels, and despite numerous attempts, sequencing reactions for the plasmids failed. We did sequence the fragment used to create pCF62 which had no mutations. We also sequenced strain AH10281 and confirmed that no mutations were present in the *yeeK* region. Plasmids pCF62 and pCF65 were not used again.

Construction of the ΔsafA$_{FL}$ strains. *safA*$_{F155STOP}$. pCF75 (Chapter 3), which contains *safA* flanked by its native promoter and terminator regions, was amplified with primers safA652D and safA684R, creating pCF124. pCF124, which holds *safA* with codon 155 exchanged from Phe to a STOP codon, was introduced in at the *amyE* locus of strain AH10357, leading to the creation of strain AH10506, where only the full length form of SafA is absent, while C30 and N21 are still produced.

safA$_{F155STOP}$-*yfp*. Primers safA-fl3 dir and fl3-yfp rev were hybridized and amplified to generate a '*safA-fl3-yfp*' fragment that contains the last 22 bp of the coding region of *safA* (minus the STOP codon), followed by the *fl3* linker (which codifies for the LGGGGSGGGGSGGGGSAAA sequence), and the beginning of the coding region of *yfp* (minus the codon for Met1). *yfp* was amplified from pKL183 (24) using primers fl3-yfp D and yfp-TsafA R which creates a fragment containing the *yfp* gene (without the codon for Met1), flanked by the end of the *fl3* linker (5' end), and the downstream region of *safA* (3' end). The two fragments were joined by overlapping PCR using primers safA-fl3 D and yfp-TsafA generating a fragment that contains *fl3-yfp* flanked by the end of the coding region of *safA* (5' end) and the downstream region of *safA*, just after its STOP codon (3' end). This last fragment was used as a megaprimer (23) to amplify pCF75 generating pCF149, which contains *safA-fl3-yfp* (referred from now on as *safA-yfp*). Finally, pCF149 was digested with *Hind*III and *Bam*HI and the released fragment was cloned in the same sites of pCF124, resulting in pCF175, which contains *safA*$_{F155STOP}$-*yfp*. For the creation of strains expressing both *safA*$_{F155STOP}$-*yfp* and *tgl*$_{wt/mut}$-*cfp* please refer to Table 4.1.

Table 4.1. Bacterial strains used in this study.

Strain	Genotype and phenotype	Origin/Construction
B. subtilis		
MB24	*trpC2 metC3*, wild type	Laboratory Stock
AH2203	*trpC2 metC3* Δ*gerQ::cm,* Cm^R	(25)
AH2255	*trpC2 metC3* Δ*tgl::sp,* Sp^R	(25)
AH10331	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35), Sp^R, Erm^R	Chapter 3
AH10281	*trpC2 metC3* Δ*yeeK::cm*, Cm^R	MB24 x pCF65
AH10282	*trpC2 metC3* Δ*yeeK::cm* Δ*tgl::sp*, Sp^R, Cm^R	AH2255 x AH10281
AH10352	*trpC2 metC3* Δ*tgl::sp* Δ*safA* Δa*myE::cm*, Sp^R, Cm^R	Chapter 3
AH10354	*trpC2 metC3* Δ*yeeK::tet*, Tet^R	AH10281 x pCm::Tc
AH10355	*trpC2 metC3* Δ*gerQ::neo*, Neo^R	AH2203 x pCm::Nm
AH10357	*trpC2 metC3* Δ*tgl::sp* Δ*thrc::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::cm* Sp^R, Erm^R, Cm^R	Chapter 3
AH10361	*trpC2 metC3 Δtgl::sp Δthrc::cm*, Sp^R, Cm^R	Chapter 3
AH10366	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{F69A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF102
AH10367	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{E115A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF104
AH10370	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{E187A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF105
AH10373	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{N188A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF107
AH10384	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{M161/164A}$, Sp^R, Erm^R, Neo^R	AH10357 x pCF108
AH10393	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{C116A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF114
AH10394	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{H200A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF115
AH10395	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{W149A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF116
AH10399	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::cm*, Δ*gerQ::neo*, Sp^R, Erm^R, Cm^R, Neo^R	AH10357 x AH10355
AH10401	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::cm*, Δ*gerQ::neo*, Δ*yeeK::tet*, Sp^R, Erm^R, Cm^R, Neo^R, Tet^R	AH10399 x AH10354
AH10416	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{W184A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF121
AH10417	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{R185A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF122
AH10418	*trpC2 metC3 Δtgl::sp Δthrc::tgl*$_{Y171A}$*-hl4-cfp*(-35), Sp^R, Erm^R	AH10361 x pCF123
AH10506	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$, Sp^R, Erm^R, Neo^R	AH10357 x pCF124

Table 4.1. Bacterial strains used in this study (*continued*).

Strain	Genotype and phenotype	Origin/Construction
AH10508	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*gerQ::cm*, Sp^R, Erm^R, Cm^R	AH10331 x AH2203
AH10509	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*yeeK::cm*, Sp^R, Erm^R, Cm^R	AH10282 x pCF89
AH10573	*trpC2 metC3* Δ*tgl::sp* Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Neo^R	AH10352 x pCF175
AH10574	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl-hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10357 x pCF175
AH10576	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::cm* Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10573 x pCF95
AH10577	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl*$_{F69A}$-*hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10576 x pCF102
AH10578	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl*$_{N188A}$-*hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10576 x pCF107
AH10579	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl*$_{C116A}$-*hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10576 x pCF114
AH10580	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl*$_{H200A}$-*hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10576 x pCF115
AH10581	*trpC2 metC3* Δ*tgl::sp* Δ*thrC::tgl*$_{Y171A}$-*hl4-cfp*(-35) Δ*safA* Δa*myE::safA*$_{F155STOP}$-*yfp*, Sp^R, Erm^R, Neo^R	AH10576 x pCF123
E. coli		
AH4611	*E. coli* BL21(DE3) with pLOM4, Kan^R	Chapter 2
AH4612	*E. coli* BL21(DE3) with pCF1, Kan^R	Chapter 2
AH4613	*E. coli* BL21(DE3) with pCF8, Kan^R	This study
AH4615	*E. coli* BL21(DE3) with pCF9, Kan^R	This study
AH4616	*E. coli* BL21(DE3) with pCF2, Kan^R	Chapter 2
AH10110	*E. coli* BL21(DE3) with pCF25, Kan^R	This study
AH10111	*E. coli* BL21(DE3) with pCF20, Kan^R	This study
AH10112	*E. coli* BL21(DE3) with pCF21, Kan^R	This study
AH10154	*E. coli* C43 with pCF19, Kan^R	This study
AH10211	*E. coli* BL21(DE3) with pCF43, Kan^R	Chapter 2
AH10325	*E. coli* BL21(DE3) with pCF68, Kan^R	This study

Antibiotic selection stands for: Cm, chloramphenicol; Erm, erythromycin; Kan, kanamycin; Neo, neomycin; Sp, spectinomycin; Tet, tetracycline.

Overexpression and purification of wild type and mutant forms of Tgl. Point mutations were introduced in the *tgl* gene present in pLOM4 (25), a pET30a(+) (Novagen) derivative where *tgl* is fused to a C-terminal His-tag, and its expression is under the control of the T7 promoter. The following set of primers were used to introduce alanine substitutions (F69A, W149A, Y171A, W184A, R185A and N188A) in *tgl*: Tgl F69Adir and Tgl F69Arev, Tgl W149dir and Tgl W149rev, Tgl Y171Adir and Tgl Y171Arev, Tgl W184Adir and Tgl W184Arev, Tgl R185Adir and Tgl R185Arev, and Tgl N188Adir and Tgl N188Arev, creating plasmids pCF19, pCF8, pCF25, pCF20, pCF21 and pCF9, respectively. These plasmids were introduced in *E. coli* BL21(DE3) (New England Biolabs) for the production of Tgl with point mutations. The only exception was pCF19 which was introduced in *E. coli* C43 (26), as the expression of Tgl^{F69A} in *E. coli* BL21(DE3) caused cell lysis. The different forms of Tgl used in this study were overproduced using the auto-induction method and subsequently purified using 1 ml columns packed with Ni-NTA Agarose resin (Qiagen), and dialyzed against 0.1 M Tris-HCl (pH 8.0), as described in Chapter 2.

***Introduction of* tgl$_{mut}$-cfp *at the* thrC *locus*.** pCF89 (which contains *tgl-cfp* for integration at the *thrC* locus, Chapter 3) was amplified with primers tgl+288D and tgl+453R, eliminating base pairs 313 to 594 of the coding region of *tgl*, yielding pCF100. '*tgl*$_{E115A}$ and '*tgl*$_{E187A}$ were obtained by the digestion of pCF43, or pCF12, respectively, with *Sfo*I and *EcoR*I, and the fragmentes were cloned in the same sites of pCF100, creating plasmids pCF104 and, pCF105. Primers Tgl F69Adir and tgl+735D were used to amplify '*tgl*$_{C116A}$, '*tgl*$_{H200A}$, '*tgl*$_{W149A}$, '*tgl*$_{N188A}$, '*tgl*$_{W184A}$, '*tgl*$_{R185A}$ and '*tgl*$_{Y171A}$ from plasmids pCF1, pCF2, pCF8, pCF9, pCF20, pCF21, and pCF25, respectively. The fragments were digested with *Sfo*I and *EcoR*I and cloned in the same sites of pCF100, creating plamids pCF114, pCF115, pCF116, pCF107, pCF121, pCF122, and pCF123. The cloning of the different fragments in pCF100 restores the *tgl* gene. Finally, primers Tgl F69Adir and Tgl F69Arev were used to amplify pCF89 which exchanges codon 69 of *tgl* from Phe to Ala, yeilding pCF102.

Table 4.2. Plasmids used in this study

Plasmid	Relevant genotype	Origin
pCF1	*tgl*$_{C116A}$-*his*$_6$, for overproduction of Tgl^{C116A}	Chapter 2
pCF2	*tgl*$_{H200A}$-*his*$_6$, for overproduction of Tgl^{H200A}	Chapter 2
pCF8	*tgl*$_{W149A}$-*his*$_6$, for overproduction of Tgl^{W149A}	This study
pCF9	*tgl*$_{N188A}$-*his*$_6$, for overproduction of Tgl^{N188A}	This study
pCF12	*tgl*$_{E187A}$-*his*$_6$, for overproduction of Tgl^{E187A}	Chapter 2
pCF19	*tgl*$_{F69A}$-*his*$_6$, for overproduction of Tgl^{F69A}	This study
pCF20	*tgl*$_{W184A}$-*his*$_6$, for overproduction of Tgl^{W184A}	This study
pCF21	*tgl*$_{R185A}$-*his*$_6$, for overproduction of Tgl^{R185A}	This study
pCF25	*tgl*$_{Y171A}$-*his*$_6$, for overproduction of Tgl^{Y171A}	This study
pCF43	*tgl*$_{E115A}$-*his*$_6$, for overproduction of Tgl^{E115A}	Chapter 2
pCF62	*yeeK* in pUC18	This study
pCF65	*yeeK::cm*, derived from pCF62	This study
pCF68	*safA*$_{C30}$-*his*$_6$, for overproduction of C30	This study
pCF75	*safA*, for integration at *amyE*	Chapter 3
pCF89	*tgl-hl4-cfp* (-35), for integration at *thrC*	Chapter 3
pCF95	*thrC::cm*	Chapter 3
pCF100	*tgl*$_{\Delta 313\text{-}594bp}$-*hl4-cfp*, derived from pCF89	This study
pCF102	*tgl*$_{F69A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF104	*tgl*$_{E115A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF105	*tgl*$_{E187A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF107	*tgl*$_{N188A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF108	*safA*$_{M161/164A}$, for integration at *amyE*	Chapter3
pCF114	*tgl*$_{C116A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF115	*tgl*$_{H200A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF116	*tgl*$_{W149A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF121	*tgl*$_{W184A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF122	*tgl*$_{R185A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF123	*tgl*$_{Y171A}$-*hl4-cfp* (-35), for integration at *thrC*	This study
pCF124	*safA*$_{F155STOP}$, for integration at *amyE*	This study
pCF149	*safA-yfp*, for integration at *amyE*	This study
pCF175	*safA*$_{F155STOP}$-*yfp*, for integration at *amyE*	This study
pCm::Nm	*cm::neo*	(24)
pCm::Tc	*cm::tet*	(24)
pET28a(+)	T7 expression vector, C-terminal His-tag	Novagen
pKL183	*yfp*	(27)

Table 4.2. Plasmids used in this study (*continued*).

Plasmid	Relevant genotype	Origin
pLOM4	*tgl-his*$_6$, for overproduction of Tgl	(25)
pMS38	*cm*	(22)
pUC18	Cloning vector	New England Biolabs

Fluorescence microscopy. Sporulation was induced by the resuspension method as described in (28). Briefly, cultures were grown in growth medium, and the cells collected at an OD_{600} of ~ 0.4, and resuspended in resuspension medium which triggers sporulation initiation. 6 and 8 hr after the beginning of sporulation, samples were taken and the cells collected by centrifugation (1 min at 2,400 x g, room temperature), and washed with 1 ml of phosphate-buffered saline (PBS). Finally, the cells were resuspended in ~20 µl of PBS and applied to microscopy slides coated with agarose (1.7%). Images were taken with standard phase contrast, CFP and YFP filters, using a Leica DM 6000B microscope equipped with an aniXon+EM camera (Andor Technologies), and driven by Metamorph software (Meta Imaging series 7.7, Molecular Devices).

Image analysis was conducted with Meta Imaging series 7.7 software and the determination of the values of fluorescence in the forespore or cell was conducted as described in Chapter 3. The individual values of the forespore/cell fluorescence ratio, (F_{FS}/F_{cell}), or cell fluorescence, F_{cell}, were examined with GraphPad Prism 5 (GraphPad Software, Inc) which indicated that the values showed a non-normal distribution. Thus, for normalization purposes each individual value was divided by the median value of the wild type strain that had been grown on the same day and under the same conditions; the same normalization was conducted for the individual values of the wild type strain. The calculation of the normalized values of the forespore/cell fluorescence ratio, or cell fluorescence, is summarized in equations 1 and 2, respectively.

(*eq. 1*) $$(F_{FS}/F_{cell})_{norm} = \frac{(F_{FS}/F_{cell})}{\text{median } (F_{FS}/F_{cell})_{wt}}$$

(*eq. 2*) $$(F_{cell})_{norm} = \frac{F_{cell}}{\text{median } (F_{cell})_{wt}}$$

Table 4.3. Oligonucleotide primers used in this study

Name	sequence (5' to 3')
yeeK-409D	GAAATCCTGATGGAAGCTTATGTTGATGTACG
yeeK+828R	GACACCCTCGAGAAAAATATCACGAATGGCTTC
yeeK+260 cat-296 D	CTCGGATAGCTTCCAGTTCCCGATATCCTGCAGGCAATAGTTAC
yeeK+170 cat+125 R	CCCAGATCACGGCATACATGGCGAGGAATTCAGAAAAAGAAG
safA652D	GCAAGACCAAT**AA**CCACAACAGGAGGCTATG
safA684R	CATAGCCTCCTGTTGTGG**TT**ATTGGTCTTGC
safA-fl3 dir	CCGTCCGGAAGAAGAAAATGAG***CTGGGCGGAGGCGGATCAGGCGGAGGCGGATCTGGCGGAGGCGGATCAGC***
fl3-yfp rev	GTTCTTCTCCTTTACT***TGCCGCAGCTGATCCGCCTCCGCCAGATCCGCCTCCGCCTGATCCGCCTCCGCCCAG***
fl3-yfp D	CTGCGGCAAGTAAAGGAGAAGAACTTTTCACTG
yfp-TsafA R	GATTTACATCGTTCCGAACGATCATTTGTATAGTTCATCCATGCCATGTG
safA+477D	GGCTATG**CCATGG**ATGGAAAATGCAAATTATCC
safA+1176R	CGTTCCGAAC**TCGAG**CTCATTTTCTTCTTCCGG
Tgl F69Adir	GCACAAAAGCGGGGCGAAG**GCT**GCCACTTTTTTAAAAACATACGGG
Tgl F69Arev	CCCGTATGTTTTTAAAAAAGTGGC**AGC**CTTCGCCCCGCTTTTGTGC
Tgl W149dir	AGAATTATTTTATATGAC**GCG**CATTATGAGAAATTGCCG
Tgl W149rev	TATCGGCAATTTCTCATAATG**CGC**GTCATATAAAATAATTCTGTC
Tgl Y171Adir	CCTTGGAGATTGTTTG**GCT**TTTAAGAATCCTGAATTTG
Tgl Y171Arev	CAAATTCAGGATTCTTAAA**AGC**CAAACAATCTCCAAGG
Tgl W184Adir	CCGCAAAAGGCGCAA**GCG**AGAGGCGAAAATGTGATAC
Tgl W184Arev	GTATCACATTTTCGCCTCT**CGC**TTGCGCCTTTTGCGG
Tgl R185Adir	CCGCAAAAGGCGCAATGG**GCA**GGCGAAAATGTGATACTACTG
Tgl R185Arev	CAGTAGTATCACATTTTCGCC**TGC**CCATTGCGCCTTTTGCGG
Tgl N188Adir	GGCGCAATGGAGAGGCGAA**GCT**GTGATACTACTGGGGGAAG
Tgl N188Arev	CTTCCCCCAGTAGTATCAC**AGC**TTCGCCTCTCCATTGCGCC
tgl+288D	GCCGCCTTCAAAAGCGATTCGGGACGCCCATGGTCTTGGAATC
tgl+453R	CCGTTTAAGATTCCAAGACCATGGGCGTCCCGAATCGCTTTTG
tgl+735R	GCGGACGATGCGGAAAAGAGACGGAACATCCAG

Underlined sequences represent introduced restriction sites; point mutations introduced are represented in **bold**; the sequence of the *fl3* linker is shown in ***bold italic***.

Overproduction and purification of C30. The coding region of C30 was amplified from genomic DNA of MB24, with primers safA+477D and safA+1176R. The resulting fragment was digested with *Nco*I and *Xho*I and introduced in the same sites of pET28a(+) (Novagen), which created pCF68,

containing *C30* fused to a C-terminal His-tag and whose expression is under the control of the T7 promoter, and is induced by IPTG or lactose. For production of C30, pCF68 was introduced in *E. coli* BL21(DE3) and the strain was grown at 37° C, for ~18 hr, in a modified auto-induction medium supplemented with 100 µg/ml of kanamycin (as described for Tgl, in Chapter 2). The cells were harvested by centrifugation (5 min at 15,300 x g, 4° C) and resuspended in 1/10 of the culture volume with lysis buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 10 mM Imidazole, pH 8.0). Cells were disrupted by 2 passages in a French Press cell (19,000 lb/in^2). Unbroken cells and cellular debris were cleared by centrifugation (30 min at 27,200 x g, 4° C). The clear lysate was applied to 1 ml columns packed with Ni-NTA Agarose resin (Qiagen) which had been previously equilibrated with lysis buffer. The column was washed with 10 ml of three different buffers: wash 1, lysis buffer containing 10% glycerol; wash 2, lysis buffer with 60 mM imidazole; and wash 3, lysis buffer with 100 mM imidazole. Finally, C30 was eluted with lysis buffer containing 250 mM imidazole. The purified fractions of C30 where dialyzed overnight against 0.1 M Tris-HCl, pH 8.0 (using SnakeSkin with a cutoff of 10 kDa, from Pierce).

Dansylcadaverine labeling assays. The assays were conducted at 50° C in Tris-HCl 0.1 M, pH 8.0, as described in Chapter 2. The following concentrations were used: BSA (New England Biolabs), 60 µM; dansylcadaverine (Fluka), 0.5 mM; and $Tgl^{wt/mut}$ (16 µM). BSA/Tgl labeling by dansylcadaverine was detected after SDS-PAGE resolution using a UV transilluminator (Chemidoc XRS, Biorad) and quantified using *ImageJ 1.37v* (29). The values obtained for each time point were normalized by the fluorescence detected for the wild type enzyme after 120 min of incubation. All Tgl forms were purified and assayed at least three times independently.

C30 cross-linking assays. Tgl (wild type or mutant forms, at a concentration of 12.5 µM) was incubated with C30 (25 µM), at 37° C, in Tris-HCl 0.1 M, pH 8.0. Samples were taken at different times and resolved in 10% SDS-PAGE gels, which were subsequently stained with Coomassie brilliant blue.

RESULTS

The substrates of Tgl appear to influence the recruitment and assembly of Tgl-CFP to the forespore. In previous work (Chapter 3), we have shown that *safA* expression is important, but not determinant, for the coat localization of Tgl. As SafA is a substrate of Tgl, we wondered if other substrates of the enzyme would also be involved in its recruitment to the coat. On the basis of their increased extractability from *tgl* mutant spores, four Tgl substrates have been proposed (30-32): YeeK, GerQ, SafA, and $SafA_{C30}$ (C30) (although additional unknown proteins may also be substrates of the enzyme; see Chapter 3). C30 corresponds to the C-terminal part of SafA, and is formed during sporulation due to an alternative site of translation initiation at codon 164 of the *safA* mRNA (Chapter 3, (33)). Concomitant with the formation of C30 is the production of $SafA_{N21}$ (N21), which corresponds to the N-terminal part of SafA (33,34). Unlike SafA and C30, there is no evidence that N21 is a substrate of Tgl (30).

To test the effects of the absence of Tgl substrates on the assembly of the enzyme, we analyzed the localization pattern of a functional Tgl-CFP fusion protein (see Chapter 3) in different mutant strains. The main period of *tgl* expression is under the control of σ^K (25,30,35,36). Thus, we analyzed the localization of Tgl-CFP in cells collected 6 and 8 hours after sporulation initiation (Figure 4.1), when the majority of the sporulating cells present visible phase dark or phase bright forespores. The cells were then sorted according to their morphological stage, into phase dark or phase bright sporangia, irrespective of the time at which the samples were collected, and the pattern of Tgl-CFP fluorescence analyzed (Figure 4.2A,B). As previously shown, in the wild type strain, Tgl-CFP is mainly detected in the forespore as two caps of fluorescence or as a ring around the developing spore (Figure 4.1a,b; classes I and II, Figure 4.2A,B). Throughout sporulation, Tgl-CFP exhibits a class I pattern in the majority of the population (panels A and B in Figure 4.2). Also, class I cells appears to increase as phase dark forespores become phase bright (from 61 to 68%) with the concurrent decrease of cells scored with a class II pattern of Tgl-CFP localization (from 38 to 30%). Because classes I and II correspond to ≥ 98% of the cells scored in the wild

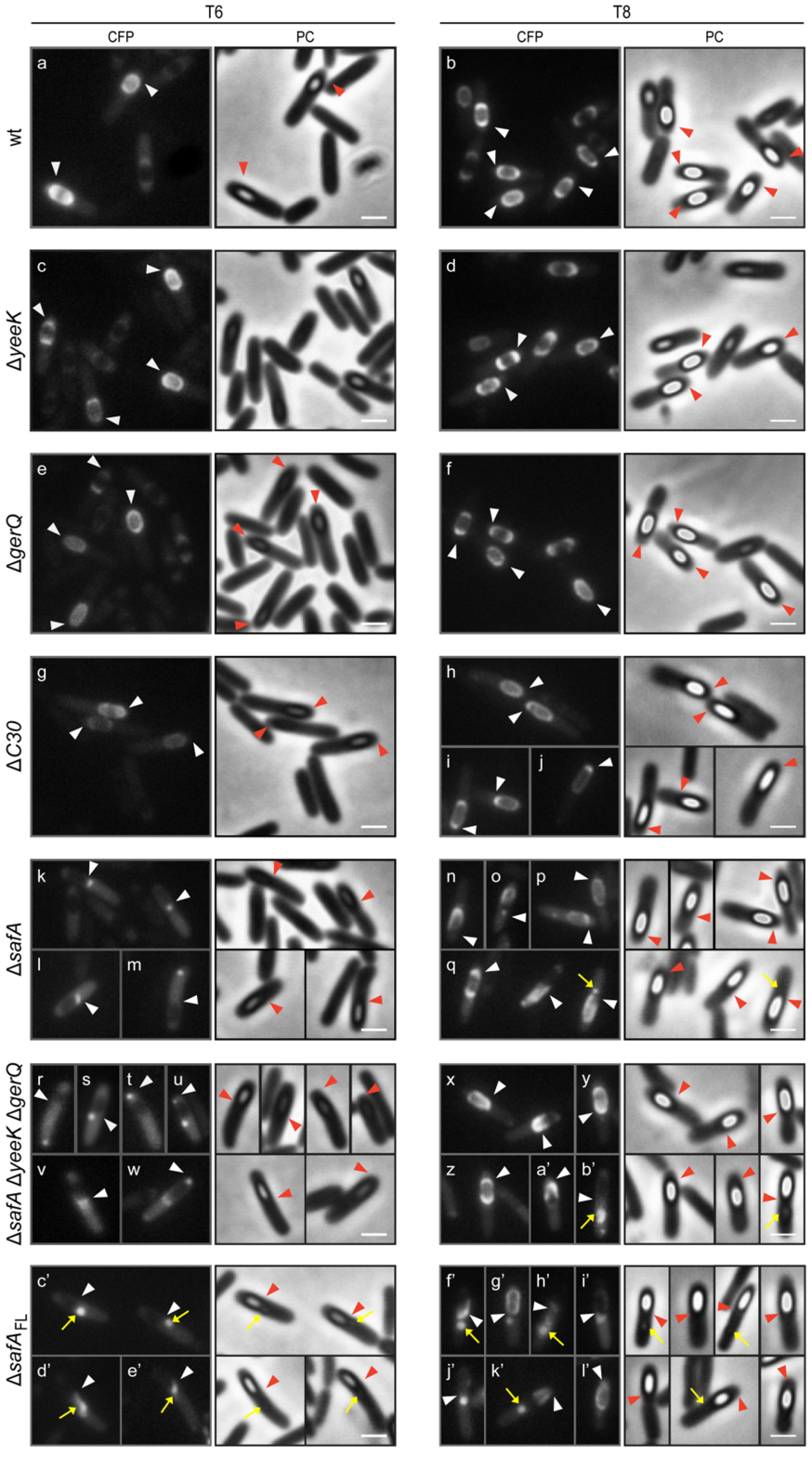
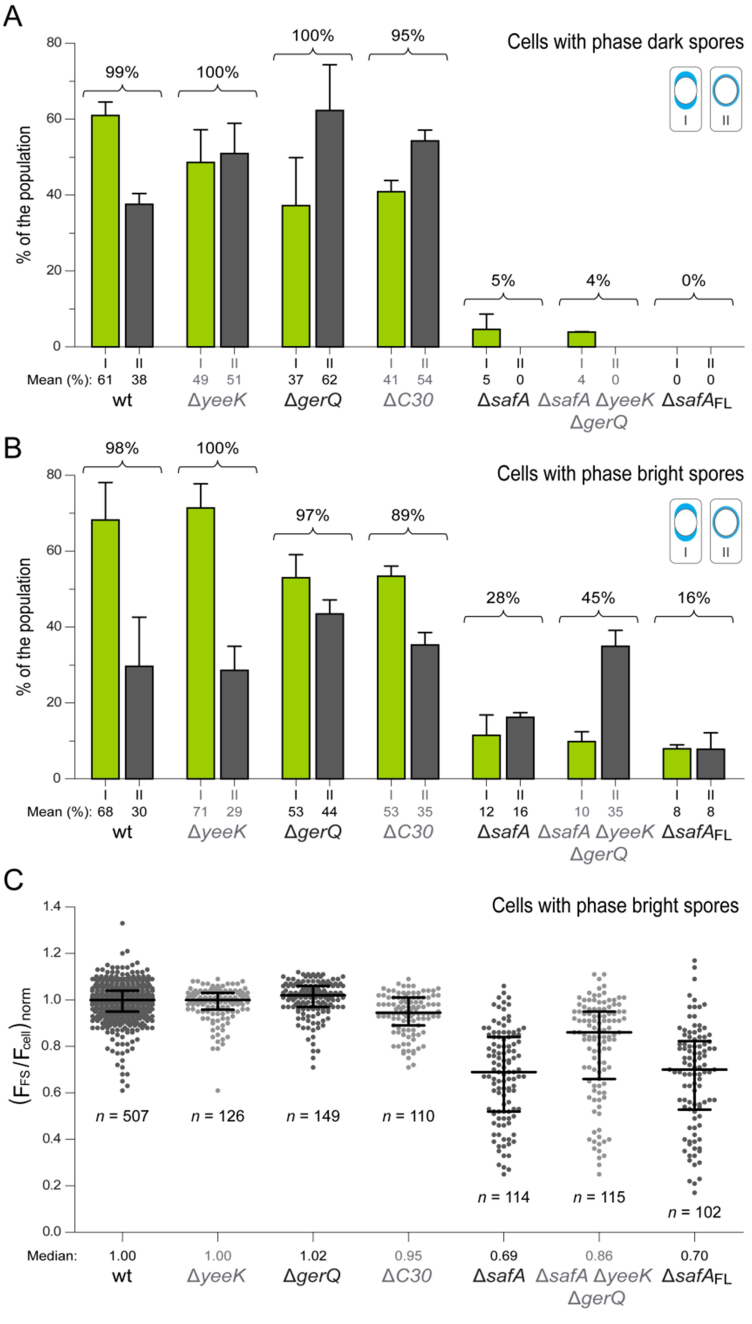

T6
T8
CFP
PC
wt
ΔyeeK
ΔgerQ
ΔC30
ΔsafA
ΔsafA ΔyeeK ΔgerQ
ΔsafA FL
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
a'
b'
c'
d'
e'
f'
g'
h'
i'
j'
k'
l'

Figure 4.1. The assembly of Tgl appears to be regulated by its substrates. *tgl-cfp* was expressed in different background strains: Δ*tgl* (wt, AH10331), Δ*tgl* Δ*yeeK* (Δ*yeeK*, AH10509), Δ*tgl* Δ*gerQ* (Δ*gerQ*, AH10508), Δ*tgl* Δ*safa safA*$_{M161/164A}$ (Δ*C30*, AH10384), Δ*tgl* Δ*safa* (Δ*safA*, AH10357), Δ*tgl* Δ*safa* Δ*yeeK* Δ*gerQ* (Δ*safa* Δ*yeeK* Δ*gerQ*, AH10401), and Δ*tgl* Δ*safa safA*$_{F155STOP}$ (Δ*safA*$_{FL}$, AH10506). Sporulation was induced by the resuspension method, and the cultures were analyzed at the microscope 6 and 8 hr after the initiation of sporulation, where the majority of the sporulating cells present phase dark or phase bright spores, as visualized by phase contrast. Representative images of the different patterns of fluorescence detected for the different strains are shown. The CFP and phase contrast (PC) channels are displayed. The arrowheads indicate the position of the mother cell proximal or distal pole of the forespore. The yellow arrows illustrate the presence of retractile material which is detected in strains were SafA is absent. White scale bar corresponds to 1 μm. See Figure 4.2 for the analysis of the localization pattern of Tgl-CFP in the different strains examined.

type strain (in both morphological stages), we considered that they represent the normal assembly pattern of Tgl-CFP. Hence, the effect of the absence of the different substrates was analyzed by determining their impact on classes I and II. Other localization patterns (see Figure 4.1) were not considered in our study.

We started by analyzing the assembly of Tgl-CFP during sporulation in the absence of YeeK, GerQ, or C30. In the majority of the cells analyzed in the *yeeK* mutant (Figure 4.1c,d), Tgl-CFP is detected in classes I and II (for both sporangia with phase dark, or phase bright spores, corresponding to ~100% of the cells scored, when compared to the ≥ 98% detected for the wild type strain; Figure 4.2A,B). In cells with phase bright spores, the number of cells scored in class I (71%) or II (29%) of the *yeeK* mutant is similar to what is seen in the wild type strain (class I: 68%, class II: 30%; Figure 4.2B). However, in phase dark sporangia of the *yeeK* mutant, about the same number of cells were scored in classes I or II (Figure 4.2A), deviating from the proportion detected in the wild type, where class I is seen in 61% of cells, compared with the 38% detected in class II (Figure 4.2A). We also quantified the forespore/cell fluorescence ratio for cells with phase bright spores to determine if the recruitment of Tgl-CFP to the spore was impaired in the absence of YeeK. As can be seen in Figure 4.2C, the fluorescence of spore-associated Tgl-CFP is similar in both the *yeeK* mutant and wild type strains (with both strains displaying a median value of 1.00). Thus, these results indicate that YeeK influences the localization of Tgl-CFP to the forespore during sporulation, although the effects of its absence are only detected in cells with phase dark spores.



In the absence of GerQ (Figure 4.1e,f), Tgl-CFP is detected in the majority of the population in classes I and II (for 100 or 97% of the cells analyzed with phase dark or phase bright spores, respectively; similar to the 99 or 98% detected in the wild type, Figure 4.2A,B). Nonetheless, the

Figure 4.2. The assembly of Tgl appears to be regulated by its substrates (*cont.*). A and B. Representation of the % of cells that display a normal Tgl-CFP localization pattern in the different strains analyzed (see Figure 4.1). Cells were separated into sporangia containing phase dark (panel A) or phase bright spores (panel B). In the majority of the population of the wild type strain (≥ 98%), Tgl-CFP localizes to the forespore either as 2 caps of fluorescence at both poles of the forespore (class I), or as a ring of fluorescence around the forespore (class II). A representation of the two categories is shown on the upper right corner of each panel. These two classes were considered as representing the normal Tgl-CFP localization pattern. The numbers above the bars correspond to the sum of the two classes. A minimum of 175 cells, and usually more than 200 cells, were quantified for each strain over two different independent experiments. The mean and standard deviation are represented. **C.** The forespore/cell fluorescence ratio was determined for individual cells and normalized by the median value obtained for the wild type strain, grown on the same day (for details see materials and methods section). The graph displays the individual normalized forespore/cell fluorescence ratios, as well as, the median and interquartile range, for the strains analyzed. The analysis was only conducted for cells displaying phase bright spores. *n* indicates the number of cells analyzed. The wild type values represented are from 8 independent experiments, while the values for the other strains were obtained from 2 different independent experiments.

individual analysis of classes I and II shows that class I displays a lower representation when compared to the wild type (Figure 4.2A,B), in both sporangia with phase dark (37/61% for Δ*gerQ*/wild type strains) or phase bright spores (53/68% for Δ*gerQ*/wild type). In fact, in cells with phase dark spores, class II represents the majority (62%) of the cells analyzed in the *gerQ* mutant, in clear contradiction with what is seen for the wild type strain (where class II corresponds to 38%, Figure 4.2A). Thus, GerQ also appears to contribute to the proper localization of Tgl-CFP during sporulation. However, the determination of the forespore/cell fluorescence ratio (Figure 4.2C), indicates that the absence of GerQ does not affect the recruitment of Tgl-CFP to the forespore, as the median value obtained for the *gerQ* mutant (1.02) is similar to that of the wild type (1.00).

We then looked at the effect of the absence of C30 in Tgl-CFP localization (Figure 4.1g-j). While Tgl-CFP was still detected in the majority of the population in a localization pattern corresponding to classes I and II, the scoring in these two classes decreased when compared to the wild type strain. In the *C30* mutant, only 95 or 89% of the cells, displayed Tgl-CFP in a class I and II pattern, for sporangia with phase dark or phase bright forespores, respectively (compared to the ≥ 98% detected in the wild type strain; Figure 4.2A,B). Furthermore, class I is reduced in the *C30* mutant when compared to the wild type, in both phase dark (41/61%, for Δ*C30*/wild type strains; Figure 4.2A) and phase bright sporangia (53/68%, for Δ*C30*/

wild type strains; Figure 4.2B). Also, similar to the *gerQ* mutant, in the absence of C30, class II (54%) represents the majority of the population in cells with phase dark spores (Figure 4.2A). In contrast, class I corresponds to 61% of the cells scored for the wild type, while class II is only seen in 38% of the population (Figure 4.2A). Moreover, the analysis of the forespore/cell fluorescence ratio for the *C30* mutant strain, shows that Tgl-CFP accumulates to slightly higher levels in the mother cell (with a median value of 0.95, compared to 1.00 of the wild type strain; Figure 4.2C).

Thus, our results imply that C30, GerQ, and YeeK may have a role in the assembly of Tgl during sporulation. The three substrates affect the localization pattern of Tgl-CFP, mainly according to a class I pattern, and C30 also seems to contribute to the recruitment of Tgl-CFP during sporulation (as detected by the decrease in the ratio of forespore/cell fluorescence when compared to the wild type). Nevertheless, in the absence of each of these substrates, the majority of the population still displays a normal distribution of the Tgl-CFP signal, corresponding to classes I and II (Figure 4.2A,B). Contrastingly, in the Δ*safA* mutant, unable to produce both SafA and C30 (Figure 4.1k-q), Tgl-CFP is detected in classes I and II in only 5, or 28% of the cells analyzed, with either phase dark or phase bright forespores, respectively, as opposed to ≥ 98% detected in the wild type, for both types of sporangia (Figure 4.2A,B). Also, the median value of the forespore/cell fluorescence ratio (Figure 4.2C) of the Δ*safA* strain (0.69) is considerably lower than that of the wild type (1.00), indicating that the elimination of *safA* leads to the accumulation of Tgl-CFP in the mother cell.

Despite the defects in Tgl-CFP localization in the *safA* mutant, in agreement with previous results (Chapter 3), the fusion protein is still able to localize, according to classes I and II, in a subpopulation of 28% of sporangia with phase bright spores. This implies that the recruitment of the enzyme is partially independent of SafA.

We then analyzed the localization of Tgl-CFP in the absence of all of the known substrates of Tgl (Δ*safA* Δ*yeeK* Δ*gerQ* strain, Figure 4.1r-b'). In this strain, Tgl-CFP was detected in class I (but not class II) in phase dark sporangia and in only 4% of the cells analyzed (similar to the localization pheno-

type observed in the Δ*safA* strain, Figure 4.2A). Surprisingly, in sporangia with phase bright spores, Tgl-CFP was detected in classes I and II in 45% of the cells scored (against 28% for Δ*safA*), and accordingly, the forespore/cell fluorescence ratio (Figure 4.2C) of the Δ*safA* Δ*yeeK* Δ*gerQ* strain (0.86) is significantly higher than that of the *safA* mutant (0.69). These results again strengthen the view that the assembly of Tgl can be directed in a *safA*-independent manner. Furthermore, in cells with phase bright spores of the *safA* mutant, both classes I and II are reduced when compared to the wild type (Figure 4.2B); in the triple *safA yeeK gerQ* mutant only class I appears to be affected (representing 10% of the cells scored when compared to 68% of the wild type), while the localization of Tgl-CFP in class II is detected in a similar percentage to that seen in the wild type strain (35% for the Δ*safA* Δ*yeeK* Δ*gerQ* mutant against the 30% of the wild type strain). Thus, this apparent localization of Tgl-CFP in class II looks independent of *safA* but it only becomes unmasked in the absence of all of the known substrates of Tgl. These results point to a role of all the known substrates in the assembly of Tgl, suggesting that there is redundancy in the recruitment of Tgl to the forespore and that two pathways may be at work, one of which is, at least under specific genetic conditions, SafA-independent.

Finally, we tried to dissect the role of C30 in Tgl-CFP localization. We have seen that the absence of C30 alone has an impact on Tgl-CFP assembly and that the single *safA* or simultaneous *safA yeeK gerQ* elimination (both strains where C30 is missing) have distinct effects on Tgl-CFP assembly. Thus, we constructed a strain where only SafA is absent, while C30 is still produced (Δ$safA_{FL}$; Figure 4.1c'-l') and analyzed the effect on Tgl-CFP assembly. The recruitment of Tgl-CFP to the forespore in the $safA_{FL}$ mutant appears similar to that seen in the Δ*safA* strain (as the forespore/cell ratio of fluorescence for both mutants is alike, Figure 4.2C). However, in the Δ$safA_{FL}$ strain, Tgl-CFP could not be detected in classes I or II in any of the cells with phase dark spores scored (Figure 4.2A), while only 16% of the sporangia with phase bright spores displayed these localization patterns (Figure 4.2B). On the other hand, in the *safA* mutant, 5 and 28% of the cells scored, are detected in classes I and II, of sporangia with phase dark or phase bright spores, respectively. This indicates that the localization of Tgl-

CFP is more compromised when SafA alone is absent (Δ*safA*$_{FL}$) than when both SafA and C30 are not present during sporulation (Δ*safA*). Because C30 localization is dependent on SafA (37) one possible explanation is that in the absence of SafA, C30 retains Tgl-CFP in a mislocalized pattern. These results suggest that C30, on its own, can affect the localization of Tgl-CFP during sporulation.

We also note that in the majority of the Δ*safA*$_{FL}$ cells with phase dark spores, Tgl-CFP accumulated as a dot of fluorescence in the mother cell proximal pole of the forespore (Figure 4.1c') (NB: this pattern will be important in subsequent sections).

Overall, our results indicate that the substrates of Tgl drive its assembly, and that two pathways may be involved. The SafA-independent pathway may nonetheless depend on other, yet unknown, Tgl substrates, whose assembly is not driven by SafA. This redundancy in the recruitment of Tgl-CFP to the forespore may explain why the assembly of the enzyme cannot be completely abolished in any of the strains analyzed.

Residues at the entrance of the tunnel of Tgl are important for the overall activity of the enzyme. If the assembly of Tgl is, as the results in the preceding section suggest, largely controlled by proteins that are also Tgl substrates, then, it should be possible to identify residues in Tgl, involved in enzyme-substrate interactions, which would also be relevant for the assembly of the enzyme. In addition, mutations in those residues of Tgl could possibly have more pronounced effects on the assembly of the enzyme than the absence of the known redundant substrates.

Tgl has a distinctive catalytic diad composed by Cys116 and Glu187, which can be non-reciprocally substituted by Glu115 (Chapter 2). The catalytic residues of Tgl are located in the middle of a tunnel that traverses the molecule from side to side and are not easily accessible from the outside (Figure 4.3). Thus, both entrances of the tunnel may act as docking ports for the substrates and residues there located may be important for substrate interactions. This in fact appears to be the case, as His200, which is located at one of the entrances of the tunnel (in the so-called back side of

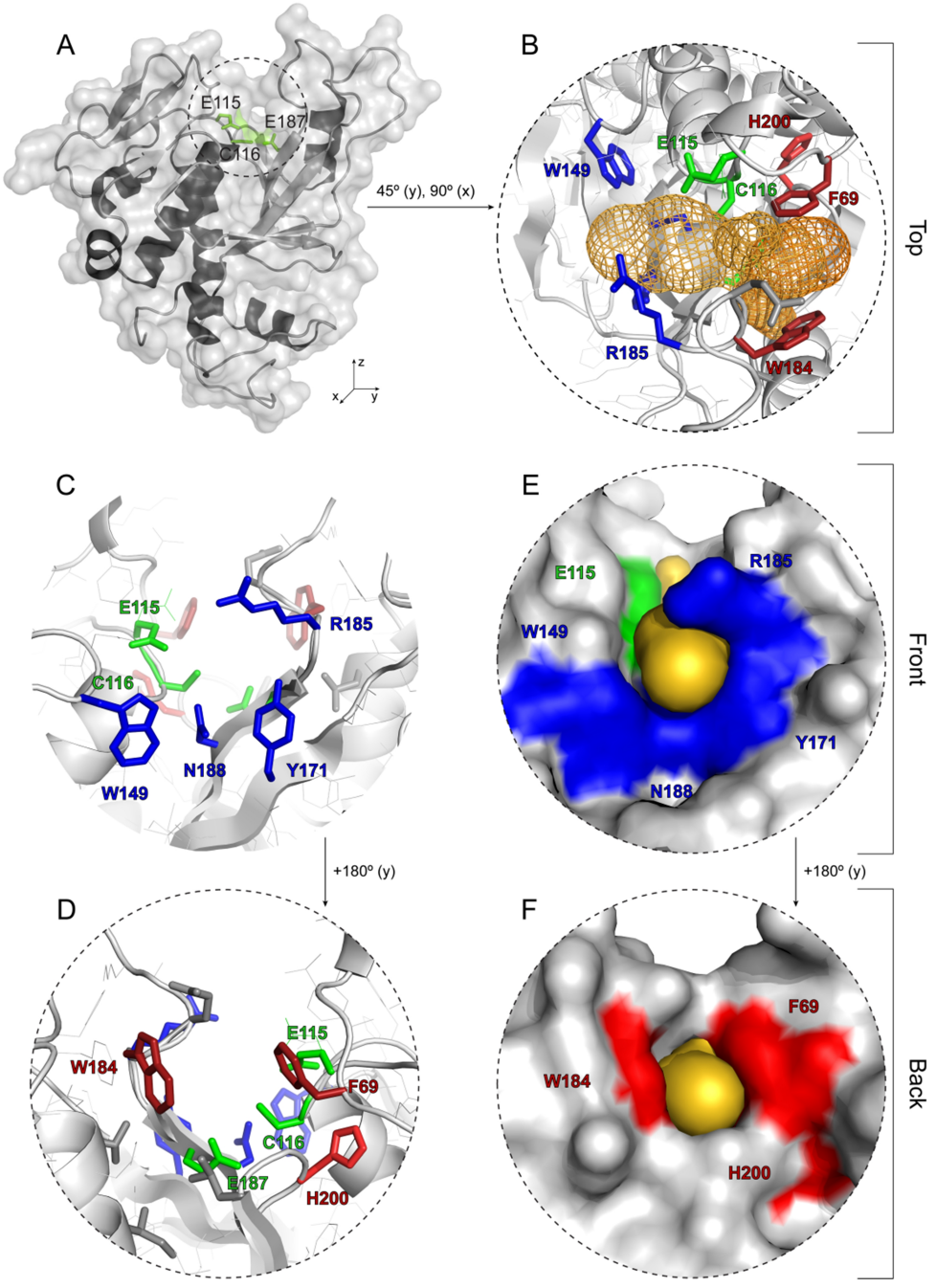


Figure 4.3. Identification of tunnel entrance residues that may be important for substrate interactions. A. Cartoon and molecular surface representation of the structure of Tgl. The side chains and surface occupancy of the catalytic residues are shown in green. It is clear that the active site of Tgl sits in a tunnel that traverses the molecule from side to side. **B-F.** Zoomed view of the tunnel region observed from the top (B), front (C,E) or back (D,F). In B-D secondary structure elements and side chains of relevant residues are shown, while in E and F only the surface is represented. Residues important within the scope of this work are highlighted in color: front, back, and catalytic residues in blue, red, and green, respectively. In B, E and F, the free space of the tunnel is represented in orange, either as a mesh (B) or by surface occupancy (E,F).

the enzyme, Figure 4.3D,F) is important for the overall activity of Tgl (Chapter 2). Additionally, because the tunnel is quite narrow, it is likely that the bulky Gln and Lys donor proteins (Q and K substrates) will approach the active site from opposite sides. This arrangement also allows for the spatial segregation of the Q and K substrates as it has been seen for animal TGases (38,39).

In search for residues, other than His200, that could also play a role in substrate interactions, we centered our attention on four residues at the front (Trp149, Tyr171, Arg185 and Asn188, Figure 4.3C,E), and two residues at the back entrance (Phe69 and Trp184, Figure 4.3D,F) of the tunnel. These residues were chosen, not only because of their position, but also because, with the exception of Arg185, they are well conserved among Tgl homologues (Chapter 2, Figure 2.5). Arg185 appeared interesting because the position of its side chain could be involved in regulating the access to the tunnel.

To test the impact of the selected residues on the overall activity of Tgl *in vitro*, each was independently substituted by Ala. The activity of the wild type and mutant forms of the enzyme was then evaluated using BSA (Bovine Serum Albumin) as the Q substrate, and the fluorescent primary amine dansylcadaverine, which substitutes for the K substrate. Enzymatic activity can thus be assessed by monitoring the formation of fluorescent BSA over time. As can be seen in Figure 4.4, all of the residues studied have a severe impact on the overall activity of Tgl. For two of the mutant enzymes tested, Tgl^{W149A} and Tgl^{Y171A}, no labeled BSA could be detected, while the exchange of residue Phe69 or Asn188 to Ala, produced enzymes with only residual activity (Figure 4.4).

The same type of assay has been previously conducted with Tgl^{E115A}, Tgl^{C116A}, Tgl^{E187A} and Tgl^{H200A} (Chapter 2, Figure 2.7). For Tgl^{E115A} and Tgl^{C116A} no activity could be detected, while Tgl^{H200A}-dependent BSA labeling was similar to what is seen for Tgl^{F69A} and Tgl^{N188A} (Figure 4.4C). As for Tgl^{E187A}, the enzyme still retained considerable activity (about 60% when compared to wild type Tgl) as Glu115 acts as a substitute (Chapter 2).

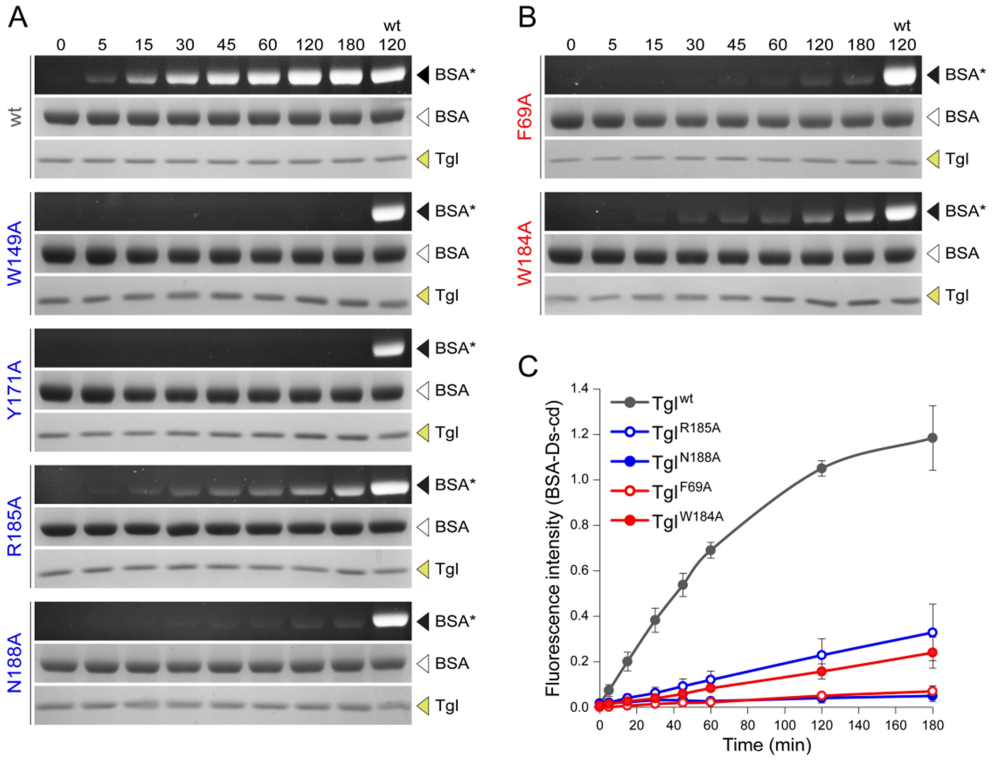


Figure 4.4. Residues at both entrances of the tunnel of Tgl are important for the enzymatic activity of the enzyme. A and B. Tgl (wild type or mutant forms) was incubated, at 50° C, with BSA and dansylcadaverine (Ds-cd), a fluorescent molecule. Samples were taken over time and resolved by SDS-PAGE. The gels were then exposed to UV light revealing labeled BSA (BSA*), and subsequently stained with Coomassie brilliant blue allowing the visualization of the total amount of BSA and Tgl loaded (BSA, Tgl). The numbers above each panel represent the time of incubation (in min) of each sample. In panel A, representative assays for the wild type (wt) enzyme and mutant forms with exchanged front side residues (see Figure 4.3) are displayed, while panel B shows the results for mutated forms of Tgl where back side residues were exchanged to Ala. **C.** To compare the activity of the different mutant forms of Tgl, the fluorescent values obtained for BSA* were normalized by the value of BSA* obtained for the wild type enzyme after 120 min of incubation (note that this sample was loaded into each gel, corresponding to the last lane). The normalized fluorescence values of BSA* were plotted for all the enzymes where fluorescence could be detected. Represented is the mean and standard error. Similar assays have been previously conducted for Tgl^{E115A}, Tgl^{C116A}, Tgl^{E187A} and Tgl^{H200A} (Chapter 2), and while for Tgl^{E115A} and Tgl^{C116A}, no activity could be detected, Tgl^{E187A} displayed about 60% of the activity of the wild type enzyme, while Tgl^{H200A} had residual activity (similar to what is seen for Tgl^{F69A} and Tgl^{N188A}).

For Tgl^{W149A}, while we could not detect the appearance of fluorescent BSA during the time of the assay (Figure 4.4A), we could, however, detect the accumulation of fluorescent Tgl^{W149A} (Figure 4.5A). The autocatalytic activity of Tgl ((40) and Chapter 2) and of other TGases has been described (41-45), and during our labeling assay, we could also detect the appearance

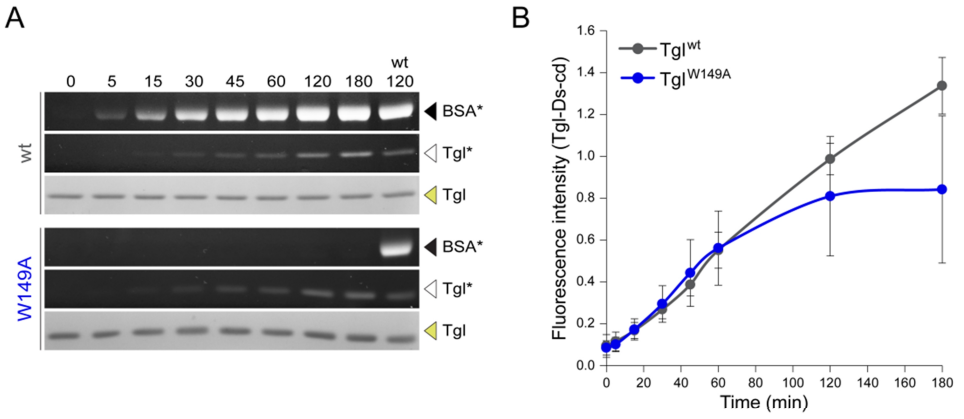


Figure 4.5. The exchange of Trp149 to Ala alters the specificity of Tgl. A. The activity of Tgl^{wt} and Tgl^{W149A} was assessed by dansylcadaverine-labeling assays as described in Figure 4.4. While no labeled BSA could be identified in the presence of Tgl^{W149A} (BSA*), Tgl^{W149A} could proficiently label itself (Tgl*). The numbers above the panel represent the time (in min) of collection of each sample. Note that the representative assay shown for Tgl^{wt} and Tgl^{W149A} is the same as that in Figure 4.4A. **B.** The fluorescent values detected for Tgl* of both enzymes were measured and normalized by the value obtained for the wild type enzyme after 120 min of incubation (a sample that was loaded into each gel, corresponding to the last lane in panel A). The normalized values were then plotted for comparison of the auto-labeling activity of both enzymes. Displayed is the mean and standard error.

of fluorescent Tgl^{wt} concurrently with the labeling of BSA (see BSA* and Tgl* in Figure 4.5A). On the other hand, Tgl^{W149A} is no longer able to use BSA as a substrate, while it can use itself. When the auto-labeling activity of Tgl^{W149A} and Tgl^{wt} is compared, both enzymes show a similar activity during the first 60 minutes of reaction (Figure 4.5B). Thus, our results indicate that Tgl^{W149A} is active, but displays an altered substrate specificity when compared to the wild type enzyme. Importantly, because BSA or Tgl are the Q substrate in the BSA/Tgl-dansylcadaverine labeling reactions, Trp149 must be involved in the interaction with the Q substrate. It follows that the front side of the tunnel, where Trp149 is located (Figure 4.3C,E), corresponds to the Q substrate docking site, while the back side of Tgl is the probable site of interaction with the K substrate (Figure 4.3D,F).

In all, these results suggest that residues at both entrances of the tunnel are involved in substrate interactions, and identify the Q and K sides of Tgl.

Front side residues of the tunnel of Tgl, likely involved in substrate interactions, are also important for the recruitment of the enzyme during sporulation. Since the substrates of Tgl seem to play an important role in

the assembly of the enzyme, Tgl mutants with Ala substitutions in residues important for substrate interactions (likely located at both entrances of the tunnel) should exhibit impaired assembly during sporulation. To test this prediction, we transferred point mutations to *tgl-cfp* inserted at the *amyE* locus of a *tgl* deletion mutant, and analyzed the assembly of the different mutant enzymes during sporulation by fluorescence microscopy (Figure 4.6). The residues analyzed included not only those located at both entrances of the tunnel, but also the catalytic residues of Tgl defined in Chapter 2 (all residues are shown in Figure 4.3).

We determined the forespore/cell fluorescence ratio of all strains producing the mutant forms of Tgl-CFP along with the wild type enzyme (Figure 4.7A). Of the four front side residues analyzed, only the exchange of Arg185 to Ala does not appear to impair the levels of the enzyme in the forespore (median values of the forespore/cell fluorescence ratio of 0.99 and 1.00, for Tgl^{R185A}-CFP and Tgl-CFP, respectively). For the other three residues, Trp149, Tyr171, and Asn188 (whose surface crafts the bottom of the front entrance of the tunnel, Figure 4.3E), their exchange to Ala severely impairs recruitment of Tgl-CFP to the forespore (median values of 0.44, 0.47 and 0.51, for the W149A, Y171A, and N188A substitutions, respectively). Moreover, these enzymes were easily detected dispersed throughout the mother cell cytoplasm (Figure 4.6e,f,h). As for the mutant forms of Tgl with exchanged catalytic and back side residues (mutations E115A, C116A, E187A, F69A, W184A, and H200A) they all seem to have their recruitment impaired to different degrees, although much less than the mutants with Ala substitutions of front residues (median values obtained for mutations on catalytic, and back side residues, are between 0.80-0.91, and 0.73-0.94, respectively; Figure 4.7A). Because the catalytic residues are not likely implicated in substrate docking, our results suggest that the activity of Tgl also contributes to its recruitment.

To test whether the recruitment defects observed for the different forms of Tgl-CFP are not simply a direct result of protein instability, the total fluorescence of the cell was measured. The individual values were then normalized using the median value of total fluorescence detected for the wild type strain that had been analyzed on the same day under the same

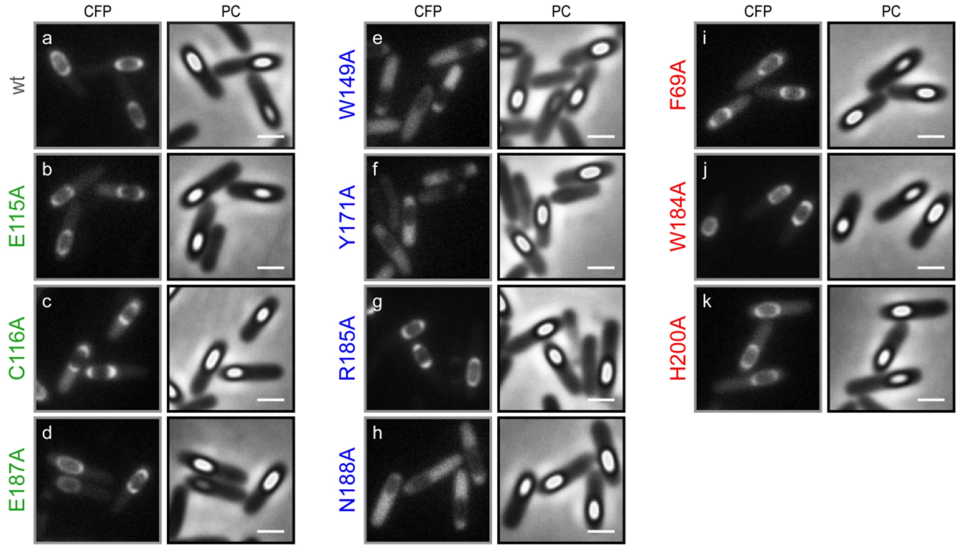


Figure 4.6. Residues located at the front entrance of the tunnel of Tgl have a defining role in the recruitment of the enzyme to the forespore. Point mutations were introduced in *tgl-cfp* creating strains expressing tgl_{E115A}-*cfp* (AH10367), tgl_{C116A}-*cfp* (AH10393), tgl_{E187A}-*cfp* (AH10370), tgl_{W149A}-*cfp* (AH10395), tgl_{Y171A}-*cfp* (AH10418), tgl_{R185A}-*cfp* (AH10417), tgl_{N188A}-*cfp* (AH10373), tgl_{F69A}-*cfp* (AH10366), tgl_{W184A}-*cfp* (AH10416), and tgl_{H200A}-*cfp* (AH10394). The assembly of the wild type (AH10331) and different mutated forms of Tgl-CFP was assessed by fluorescence microscopy. Sporulation was induced by the resuspension method and the cultures were analyzed at hour 8 of sporulation. A field representative of each strain for the CFP and phase contrast (PC) channels is shown. Residues exchanged to Ala are identified on the left side of each set of images and are color coded according to their role in activity or position within the structure of Tgl (see Figure 4.3): catalytic, front entrance, and back entrance residues indicated in green, blue, and red, respectively. Scale bar, in white, corresponds to 1 µm. For a detailed analysis of each strain see Figure 4.7.

experimental conditions (Figure 4.7B). As a control, and to ascertain the intrinsic variability of this type of measurement, the wild type strain was grown in four independent cultures on the same day and treated as if corresponding to four different strains. The total fluorescence was quantified and normalized by the culture which was arbitrarily considered as the wild type (insert in Figure 4.7B, light gray values). The results show that there is some intrinsic variability associated with this kind of measurement (median values of 0.95, 0.88, and 0.94, for a-c, in Figure 4.7B). However, some of the point mutations introduced in *tgl-cfp* lead to proteins that accumulate to significantly lower levels than when compared to the wild type enzyme and below the threshold of intrinsic variability. This may indicate that some of the mutations introduced may alter the folding of Tgl-CFP. Nonetheless, there doesn't appear to be a direct correlation between

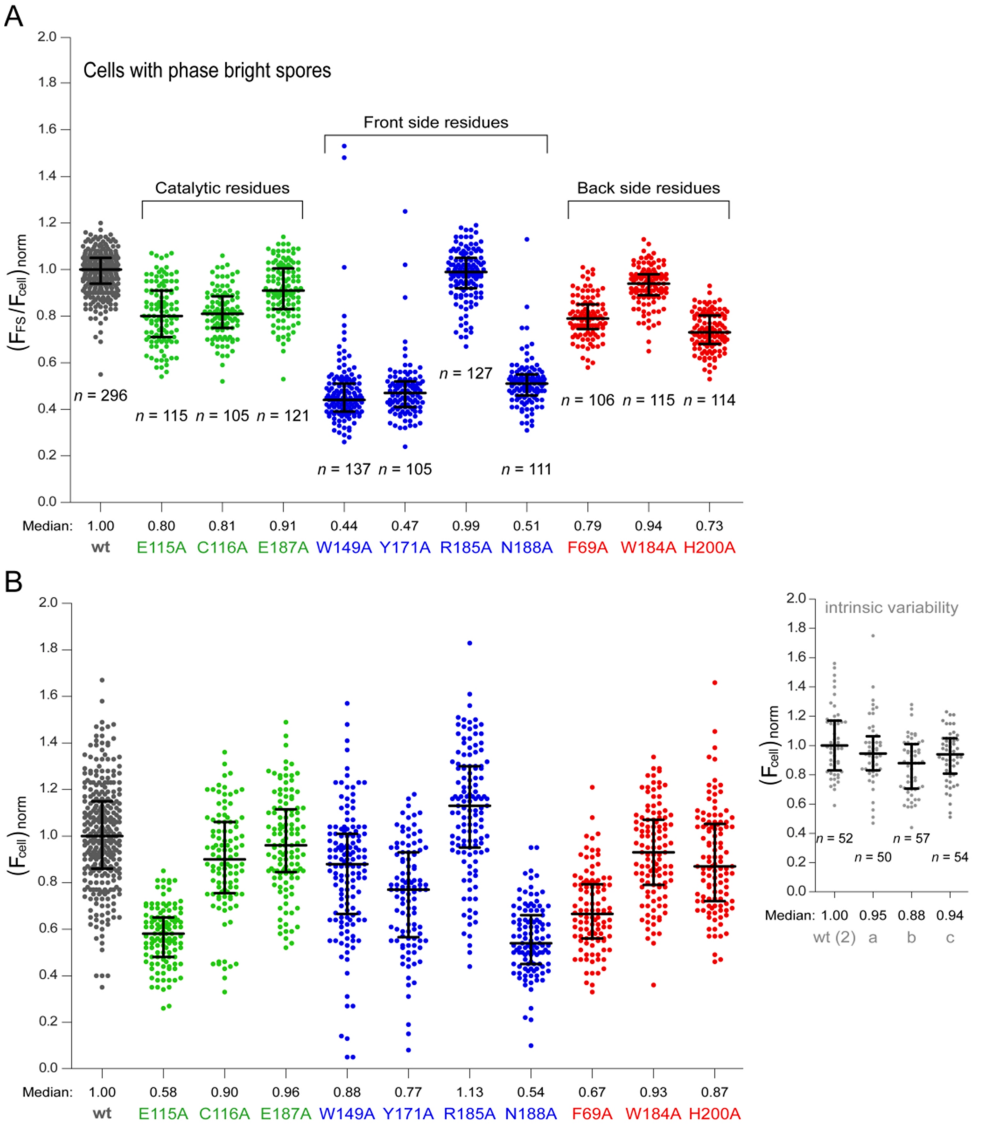


Figure 4.7. Residues located at the front entrance of the tunnel of Tgl have a defining role in the recruitment of the enzyme to the forespore (*cont.*). The effect of exchanging different residues of Tgl to Ala was investigated by fluorescence microscopy (Figure 4.6). **A.** Cells bearing phase bright spores were analyzed and the forespore/cell fluorescence ratio was measured and normalized (similar to what was done in Figure 4.2C). Displayed are the individual normalized ratio values, the median, and interquartile range, for each strain. **B.** The fluorescence signal detected in the whole cell was quantified and normalized using the median value of the wild type strain grown on the same day (see materials and methods for details). Note that the same cells were analyzed in panels A and B. The insert in panel B (light gray values, on the far right) represents the intrinsic variability associated with the type of measurement conducted: as a control, the wild type strain was grown as four independent strains [wt (2), and a-c], and the fluorescence of the cells determined and normalized by the median value of the arbitrarily chosen wild type strain [wt (2)]; the four sets of values obtained are shown. Individual normalized values, the median, and interquartile range are plotted. *n* indicates the number of cells quantified.

the levels at which the different mutant forms of Tgl-CFP accumulate in the cell and their recruitment to the forespore (e.g., Tgl^{E115A}-CFP and Tgl^{F69A}-CFP, whose fluorescence levels in the cell are among the lowest, at 0.58 and 0.67, respectively, but whose recruitment is not of the most affected; median value of forespore/cell fluorescence ratio of ~0.80, while the three minimum values are between 0.44 and 0.51). Thus, the defects observed in the recruitment of the different mutant forms of Tgl-CFP are not a direct result of protein instability and likely reflect impaired enzyme-recruiter interactions. In sum, these results suggest that the assembly of Tgl relies on enzyme-substrate interactions, with Q-side residues (located at the front entrance) making the most important contribution to the assembly of the enzyme.

Residues important for the recruitment of Tgl to the spore are also important for the* in vitro *cross-linking of C30. One projection arising from the results presented in the preceding section is that at least some of the mutant forms of Tgl should show reduced activity towards its substrates. When analyzing the assembly dependency of Tgl on its substrates, C30 appears to be the one (other than SafA) whose absence mostly impacts the assembly of the enzyme (Figure 4.2). Moreover, C30, but not the other known substrates of Tgl, could be overproduced in soluble form and purified (as detected by Coomassie staining after SDS-PAGE resolution). As such we decided to test the activity of different Tgl^{mut} proteins in the cross-linking of C30 *in vitro*. We analyzed the activity of Tgl mutant forms whose recruitment appears most compromised, both for substitutions of residues at the front and back entrances. As a control, we also tested the activity of two Tgl mutant forms with exchanged catalytic residues.

C30 and Tgl (wild type and mutant forms) were independently over-expressed in *E. coli* with a C-terminal His-tag, and the various proteins were purified and subsequently incubated together at 37° C. This is the temperature at which the fluorescence microscopy experiments were conducted; note however that the optimal temperature of Tgl is 50° C and it was at this temperature that the dansylcadaverine labeling studies were conducted (Figures 4.4 and 4.5). As can be seen in Figure 4.8A, wild type Tgl

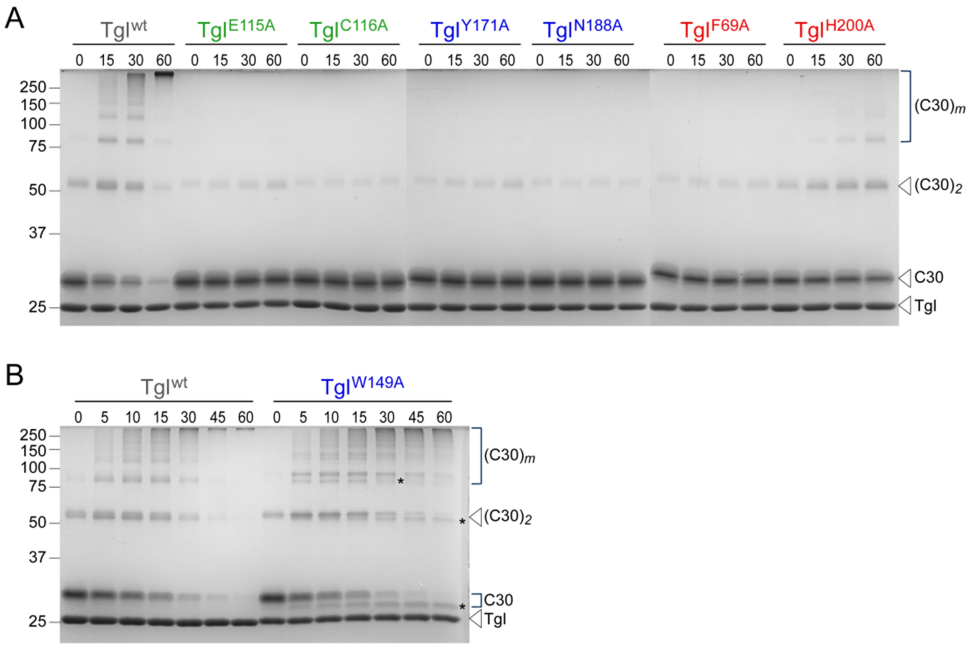


Figure 4.8. Non-catalytic residues, important for the recruitment of Tgl to the forespore during sporulation, are also important for the *in vitro* cross-linking of C30. C30 and Tgl, wild type and mutant forms, were incubated at 37° C, and samples were collected at different times and resolved by SDS-PAGE (the incubation time, in min, is displayed above each lane). The gels were stained with Coomassie brilliant blue for protein visualization; Tgl and the different forms of C30 are identified on the right hand side of each panel. $(C30)_2$ corresponds to a possible dimer of C30, while $(C30)_m$ represents the site of migration of multimeric forms of C30. The asterisks in panel B indicate the site of migration of forms of C30 that are only detected in assays with Tgl^{W149A}. Enzymes are color coded according to the location of the residues exchanged relative to the tunnel of Tgl: green, blue, and red represent catalytic (middle of the tunnel), front, and back entrance residues, respectively. The site of migration of the molecular weight markers, in kDa, is indicated on the left side of each panel.

efficiently cross-linked C30 into high molecular weight species. For Tgl^{C116A} and Tgl^{N188A}, no activity could be detected. As for Tgl^{E115A}, Tgl^{Y171A}, and Tgl^{F69A} we could detect residual activity (see the C30 species identified as $(C30)_2$, in Figure 4.8A). In contrast, Tgl^{H200A} displayed significant activity when compared to the other mutants forms of the enzyme, with the visible formation of high molecular weight products during the time of the assay (Figure 4.8A). As for Tgl^{W149A}, we had previously seen that the enzyme displays significant enzymatic activity, although its specificity appears altered (Figure 4.5). Yet, our results have also shown that the recruitment of Tgl^{W149A} to the forespore is severely compromised (Figures 4.6e and 4.7A), which appears to be a contradiction. However, our C30 cross-linking assays show that Tgl^{W149A} produces forms of C30 that are not detected with Tgl^{wt}

or any of the other mutant forms of the enzyme (asterisks in Figure 4.8B). One possibility is that these forms correspond to C30 with intramolecular cross-links. Presumably, the altered interaction of Tgl^{W149A} with C30 does not support assembly of the enzyme *in vivo*.

C30 retains Tgl^{wt} in the mother cell but not Tgl mutants with altered front side residues. As stated before, in the Δ*safA*$_{FL}$ strain, Tgl-CFP is detected in the majority of the cells with phase dark forespores as a bright dot of fluorescence in the mother cell proximal pole of the forespore (Figures 4.1c' and 4.9). This pattern was also detected in the Δ*safA* single and Δ*safA* Δ*yeeK* Δ*gerQ* triple mutant, although not as frequently (Figure 4.9). This dot of fluorescence appeared to coincide with the accumulation of refractile material (yellow arrows in Figure 4.1). One possibility we considered was that C30 could be retaining Tgl, along with other coat proteins. At later times, Tgl-CFP tends to lose this single dot localization (the representation of which decreases to 12%, Figure 4.9; note that while sometimes the fluorescence dot is still visible in the mother cell, it is not the only pattern of fluorescence detected in the same cell, thus it was not considered as corresponding to the single dot pattern; as examples see Figure 4.1f',g',k'). The presence of refractile material was also consistently detected in cells with phase bright spores, indicating again, that along with Tgl other proteins are retained in the mother cell cytoplasm of the Δ*safA*$_{FL}$ strain. We reasoned that if this interpretation is correct, than in the Δ*safA*$_{FL}$ strain Tgl and C30 should co-localize. Moreover, Tgl mutant forms whose recruitment and ability to cross-link C30 *in vitro* is impaired should be detected throughout

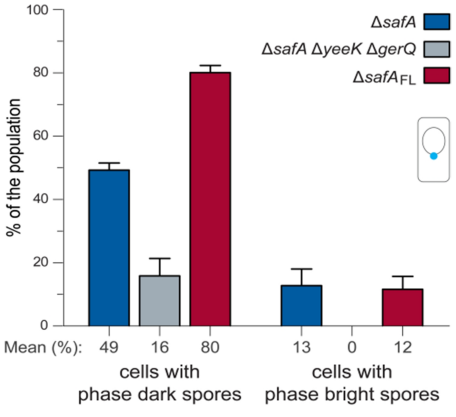


Figure 4.9. Before cortex maturation, in Δ*safA*$_{FL}$, C30 traps Tgl-CFP in an erroneous localization. Representation of the number of cells that displayed a pattern of localization as single dot of fluorescence on the mother cell proximal pole of the forespore (exemplified on the right side; see also Figure 4.1k,s,c'). Only strains where this pattern of localization was observed are illustrated. Represented is the mean and standard deviation.

the mother cell cytoplasm.

We expressed *C30-yfp* along with *tgl-cfp* in a Δ*safA* strain. Because the single fluorescence dot pattern of Tgl-CFP is only preponderant in cells with phase dark spores (Figure 4.9), for simplification purposes, the co-localization of Tgl-CFP and C30-YFP was only scored in these cells. As can be seen in Figure 4.10Aa,B, C30-YFP localizes as a bright dot of fluorescence in the majority of the population (84%), and Tgl^{wt}-CFP co-localizes with its substrate (in 76% of the cells examined). Tgl mutants with exchanged back side residues or Tgl^{C116A}-CFP, also co-localize with C30-YFP, although at a lower percentage (values varying between 69 and 70%, Figure 4.10Ab,f,g,B). For these Tgl^{mut}-CFP forms, we could also detect a higher fluorescence signal in the mother cell cytoplasm when compared to the wild type enzyme (Figure 4.10A). This is in accordance with our previous results that show that these mutant forms of the enzyme have their recruitment to the forespore slightly impaired (Figures 4.6 and 4.7A). Finally, the examination of the mutant forms of Tgl with exchanged front site residues, Tgl^{Y171A}-CFP and Tgl^{N188A}-CFP (Figure 4.10Ac,d,B), reveals that these two forms largely accumulate throughout the mother cell cytoplasm, and Tgl^{Y171A}-CFP and Tgl^{N188A}-CFP only co-localize with C30-YFP in 3 and 10% of the population, respectively. Note, however, that C30-YFP still continues to localize in these strains as a fluorescent dot at the mother cell proximal spore pole in 77 and 80% of the cells of strains carrying *tgl*$_{Y171A}$*-cfp* and *tgl*$_{N188A}$*-cfp*, respectively. Thus, it appears that these mutant forms of Tgl with exchanged front side residues, that have their recruitment to the forespore severely impaired (Figures 4.6 and 4.7A), are also not being retained in a fluorescent dot pattern by C30. This observation strongly suggests that the front side residues of the Tgl tunnel play a crucial role in the assembly of the enzyme.

DISCUSSION

This work reports a recruitment mechanism for an enzyme of the spore coat whose assembly into the forespore appears to be driven by its substrates. Our assumption that the recruitment and proper localization of Tgl is driven by its substrates is based on two observations: the absence of the sub-

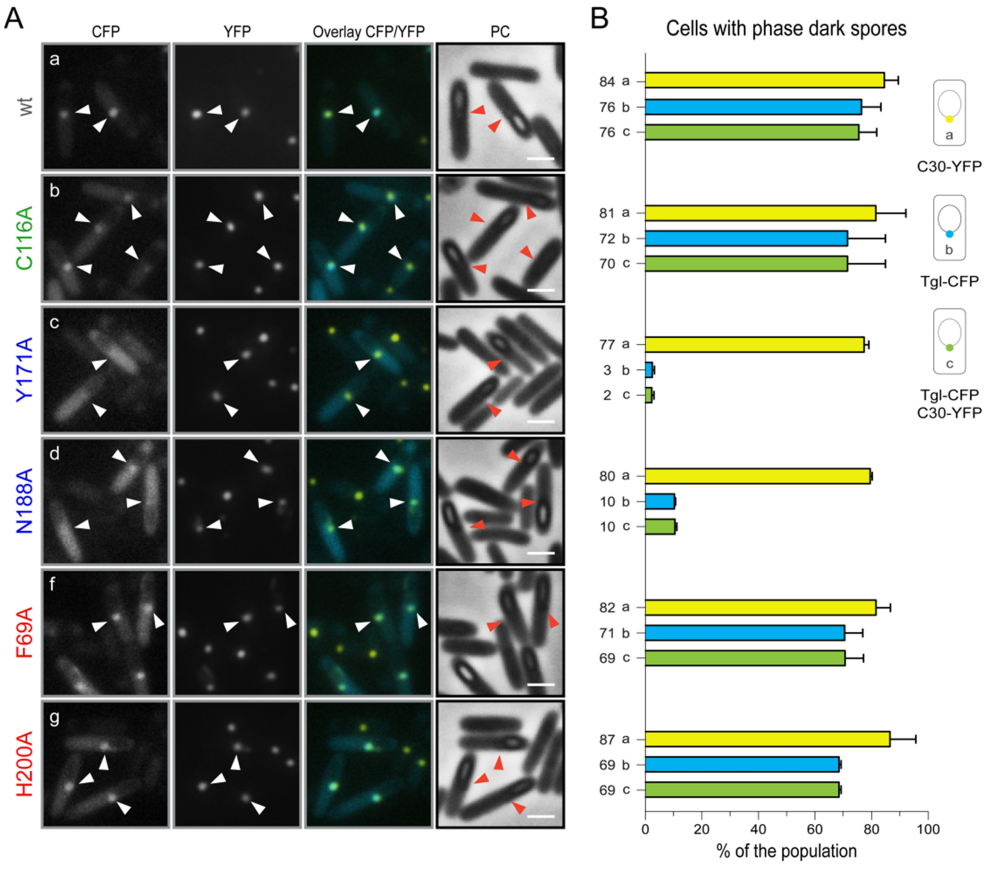


Figure 4.10. The aggregation of *tgl-cfp* seen in Δ*safA*$_{FL}$ is lost in *tgl-cfp* mutants where front side residues were exchanged to Ala. *tgl-cfp* was introduced in a Δ*safA*$_{FL}$ strain expressing *C30-yfp*. Different forms of *tgl-cfp* were analyzed in this background: *tgl*$_{wt}$*-cfp* (AH10574), *tgl*$_{C116A}$*-cfp* (AH10579), *tgl*$_{Y171A}$*-cfp* (AH10581), *tgl*$_{N188A}$*-cfp* (AH10578), *tgl*$_{F69A}$*-cfp* (AH10577) and *tgl*$_{H200A}$*-cfp* (AH10580). Cultures were grown and sporulation induced by the resuspension method. Samples were collected 6 hr after sporulation initiation and analyzed by fluorescence microscopy. Strains are identified according to the *tgl-cfp* form that they express and color coded according to the location of the exchanged Tgl residues: catalytic residues, front and back side entrances of the tunnel in green, blue and red, respectively. This identification of the strains applies to panels A and B **A.** Figures representative of images taken with CFP, YFP and phase contrast (PC) filters are displayed. The overlay between CFP and YFP images is also shown. Scale bar, in white, corresponds to 1 µm. Arrow heads indicate the position of the C30-YFP signal in cells where a phase dark forespore could be detected. **B.** Cells with phase dark spores were examined. Plotted is the % of cells that showed a C30-YFP localization as a dot of fluorescence in the mother cell proximal pole (class a) as well as the cells where Tgl-CFP displayed such a pattern (class b). Finally, the cells were also scored for the co-localization of C30-YFP and Tgl-CFP as dot of fluorescence in the mother cell proximal pole of the forespore (class c). A diagram of each class of localization can be seen on the right hand side. The graph displays the mean and standard deviation for each class. The numbers on the left side indicate the mean value obtained for each class in each strain. A minimum of 201 cells were analyzed for each strain in two sets of independent experiments.

strates affects the assembly of the enzyme to different degrees, and residues of Tgl most likely involved in substrate interactions are determinant for the deposition of the enzyme onto the surface of the developing spore. Our data also suggests that the substrates have redundant roles in the recruitment of the enzyme, and that two pathways for Tgl assembly are at work, one which is SafA-dependent, and a second which is not.

A few observations merit further discussion. First, the analysis of the assembly of Tgl-CFP in the absence of *yeeK*, *gerQ* or *C30* shows that the effects of the elimination of *yeeK* are only detected in cells with phase dark forespores (Figure 4.2A,B), while the absence of GerQ or C30 affects the localization of the enzyme to both phase dark and phase bright forespores. Because GerQ and C30 expression is under the control of σ^E (unlike YeeK which is σ^K dependent), it follows that by the time *tgl* transcription is initiated (under σ^K control), GerQ and C30 are most likely already assembled onto the forming coat. Hence, pre-assembled substrates appear to have a more important role in the localization of Tgl during sporulation.

Second, the presence of a SafA-independent pathway in the assembly of Tgl-CFP was more clearly identified in a strain were all the known substrates of Tgl where eliminated, the triple *safA yeeK gerQ* mutant. In this strain, the assembly of Tgl-CFP is not as compromised as in the single *safA* mutant (Figure 4.2). The reason for this is unknown. However, because SafA controls the assembly of inner coat proteins one explanation is that in the *safA* mutant, both YeeK and GerQ (inner coat proteins, (46)) are mislocalized and are contributing to the abnormal Tgl-CFP localization pattern. In support of this hypothesis is the observation that the localization of YeeK-GFP, in a *safA* mutant, is abnormal and the protein tends to accumulate at the mother cell proximal pole of the forespore (47). Furthermore, our own unpublished data suggests that the levels of GerQ are reduced in *safA* mutant spores (as detected by Coomassie staining), implying that the proper assembly of GerQ is reliant on *safA*. At this point we would also like to note that the dispersion of the forespore/cell fluorescence values for the *safA*, *safA yeeK gerQ*, and *safA*$_{FL}$ mutant strains, is much broader than when compared with the remaining strains analyzed (Figure 4.2C). This reflects the greater heterogeneity detected in Tgl-CFP

localization patterns which were detected in the *safA*, *safA yeeK gerQ* and *safA*$_{FL}$ mutants (Figure 4.1).

Third, the redundant role of the various Tgl substrates in the recruitment of the enzyme to the forespore may account for the less drastic defects seen in the *yeeK*, *gerQ* or *C30* mutants when compared with strains where *safA* was eliminated (where several Tgl substrates are absent or mislocalizing). Furthermore, SDS-PAGE gels of soluble coat proteins extracted from *tgl* mutant spores, suggest that all the substrates of the enzyme are more abundant than Tgl itself (as an example see Figures 3.2 and 3.4) as, unlike the enzyme, they can all be easily detected by Coomassie staining. Therefore, it is possible that the absence of one of the substrates may be compensated by the fact that the enzyme is present at lower levels than its substrates.

Fourth and final, we have uncovered a role for C30 during sporulation. In the Δ*safA*$_{FL}$ strain, by phase contrast microscopy, we consistently detected the presence of refractile material in the mother cell cytoplasm. Although the same material could be detected in the *safA* mutant, it was not as frequent as in Δ*safA*$_{FL}$ (Figure 4.1). This phenotype is reminiscent of what is seen in mutant strains of morphogenetic proteins (like for example in a *spoVID* mutant), and most likely corresponds to the aggregation of spore coat proteins. In fact, Tgl-CFP appeared to localize on top of the refractile material (especially in cells with phase dark forespores) where it co-localized with C30-YFP (Figure 4.10). Thus, C30 (whose localization is dependent on SafA and hence unable to correctly localize in the Δ*safA*$_{FL}$ strain) appears to dictate the localization of several coat proteins, trapping them in a mislocalized bundle. While N21 is also produced in the *safA*$_{FL}$ mutant (even if not in its entirety), it is unlikely that N21 is causing the aggregation of coat proteins, as a strain that only produces C30 in the absence of SafA also shows the same defects (not shown). This implies that the SafA-dependent inner coat assembly is mainly directed by the C-terminal part of SafA, and that C30 may contribute by providing additional binding sites for interaction with inner coat proteins. However, because *C30* elimination does not originate major defects in coat assembly (Figures 3.1C and 3.2B), that means that SafA can compensate for the lack of C30. In line with the view of

C30 as a SafA "helper", of all the known substrates of Tgl, C30 is the one whose absence most impairs the localization of Tgl-CFP (apart from the elimination of *safA* which likely has both direct and indirect consequences, as discussed above). As such, C30 and by extension the C-terminal part of SafA, may be the major factors in the recruitment of Tgl during sporulation.

The substrate driven assembly of Tgl is reminiscent of that observed for several penicillin binding proteins (PBPs) involved in cell wall synthesis. PBPs are distinguished according to their structural domains, but they all share a common penicillin-binding domain with D,D-peptidase activity. The substrate for D,D-peptidase activity corresponds to the peptidoglycan stem peptide, and PBPs catalyze D,D-transpeptidation, D,D-carboxypeptidation or D,D-endopeptidation reactions (48,49). Although many PBPs localize in a substrate-independent manner, in recent years it has become clear that the localization of some of these proteins is driven by substrate availability. Additionally, this mechanism is widespread across different bacterial species and examples include: PBP2 of *Staphylococcus aureus* (50), PBP5 of *Escherichia coli* (51), PBP3 of *Caulobacter crescentus* (52), and PBP2A and PbpH of *B. subtilis* (53) (among others (54)). It was shown that inhibition of the synthesis of the substrates of PBPs or the use of binding competitors for PBPs' substrates (i.e., the reduction or elimination of available substrates) leads to the delocalization of the enzymes (50,52). This is similar to what happens with Tgl in the absence of its substrates. For Tgl, however, the effects in localization are not as drastic because there is redundancy among the recruiters of the enzyme. Moreover, the induced delocalization of the substrates of PBP2A and PbpH of *B. subtilis*, leads to the delocalization of the two enzymes (53). In a clear resemblance, the accumulation of C30 at the mother cell proximal pole of the forespore in the Δ*safA*$_{FL}$ strain, also delocalizes Tgl to the same spot (Figure 4.10). Finally, the localization of PBP3, of *C. crescentus* or PBP5, of *E. coli*, is dependent on enzymatic activity, as the mutation of the active site Ser residue to Ala (PBP3) or Gly (PBP5) leads to the formation of inactive enzymes which no longer correctly localize. While catalytic residues appear important for the proper recruitment of Tgl-CFP to the forespore, they are not essential. However, unlike Tgl (see below), the activity of PBPs is not delayed following their

synthesis, thus substrate binding and catalysis are probably coupled. In accordance to this, the PBP5^{S44G} mutant loses its ability to bind penicillin, which mimics the physiological substrate of PBP5 (48).

Altogether, our results point to a substrate driven assembly of Tgl during sporulation. Furthermore, our results indicate that the front side residues of Tgl are involved in the interaction with Q substrates and are paramount for the recruitment of the enzyme to the forespore. On the other hand, back side (docking side of the K substrate) and catalytic residues, also appear to contribute (but to a considerable less extent) to the proper localization of Tgl. This implies that the interaction of catalytic and back side residues with the substrates follows, and most likely only occurs, after the substrate interacts with the front entrance of the tunnel.

The fact that catalytic (buried) residues also play a role in the assembly of the enzyme indicates that the catalytic activity of Tgl is important for its recruitment to the forespore. This is somewhat unexpected as Tgl-dependent activity is detected many hours after the assembly of the enzyme, following mother cell lysis (25,31). However, a closer look at the results shows that it is not activity *per se* that dictates the recruitment of Tgl. Cys116 is the critical catalytic residue of Tgl (Chapter 2) and, accordingly, Tgl^{C116A} shows no activity in different *in vitro* assays or *in vivo* (Chapter 2: Figures 2.7A and 2.8B; Figure 4.8A; and (25,55)). Tgl^{H200A}, on the other hand, still retains activity in all the *in vitro* activity assays (Chapter 2: Figures 2.7A and 2.8B; and Figure 4.8A), but its recruitment to the forespore is more impaired than Tgl^{C116A} (Figure 4.7A). These results suggest that three different steps are involved in the assembly of Tgl, which may be sorted according to the role of the different residues in the activity of the enzyme.

The reaction mechanism of transglutaminases follows an acylation-deacylation pathway (Figure 2.1, (56)). Initially, the catalytic Cys residue of transglutaminases reacts with the Q substrate (in Tgl this step is also dependent on Glu187 or Glu115), leading to the formation of an acyl-enzyme intermediate where the substrate and the enzyme are covalently bound. The K substrate then reacts with the acyl-enzyme intermediate leading to the formation of the cross-linked product and its release from the

enzyme. Thus, the assembly of Tgl may be seen in the following manner: first the Q substrate interacts with the front side residues of Tgl; this is followed by the formation of the acyl-enzyme intermediate which further stabilizes the Tgl-substrate complex (explaining why catalytic residues are important for the forespore recruitment of the enzyme); and finally, a second Tgl-substrate interaction is created at the back side entrance of the tunnel, with the K substrate. This second interaction may be established with the same substrate molecule that is already associated with Tgl or with a different one, which would lead to the formation of a tripartite complex. Nonetheless, because back side residues cannot compensate for the absence of front side residues, the K substrate-Tgl interaction is likely to depend on the initial Q substrate-Tgl interaction. Because Tgl remains inactive until mother cell lysis, the K substrate presumably interacts with Tgl in a conformational inadequate position, so that the Lys residue is unable to react with the acyl-enzyme intermediate.

The substrate driven assembly mechanism of Tgl appears as an elegant and simple way of controlling the localization of the enzyme. Furthermore, it ensures that Tgl is only recruited to the forespore when its substrates are present, and that Tgl is correctly localized (near its substrates) when its cross-linking activity is triggered. This most likely helps in the regulation of the activity of Tgl, keeping it in check and guarantying that the correct proteins or assemblies are cross-linked. This mechanism shows some similarities to that of Factor XIIIa (FXIIIa), the transglutaminase involved in the last stages of the blood coagulation cascade. FXIIIa is responsible for the cross-linking of fibrin (as well as other secondary proteins) during the formation of the blood clot. Factor XIII (FXIII), the zymogenic inactive form of FXIIIa, exists in the plasma as an A_2B_2 heterotetramer. The A_2 subunits correspond to the transglutaminase zymogen, while B_2 is a non-enzymatic glycoprotein, and it has been shown that the majority of FXIII present in the plasma is bound to fibrinogen by its B_2 subunits (57). Fibrinogen, in turn, corresponds to the unprocessed form of fibrin. Thus, the inactive transglutaminase is associated with its substrate precursor ensuring its proper localization for the formation of the blood clot, similar to what is seen with Tgl. FXIII activation will them be elicited by its processing by

thrombin (which also cleaves fibrinogen into fibrin) and dissociation from the B_2 subunits, by calcium binding and even by fibrin, to produce FXIIIa. For Tgl however, the mechanisms by which its activity is triggered are still unknown, but they are likely dependent on late coat maturation processes that occur after mother cell lysis.

REFERENCES

1. **Henriques AO, Moran CP, Jr.** 2007. Structure, assembly, and function of the spore surface layers. Annual review of microbiology **61:**555-588.

2. **Kim H, Hahn M, Grabowski P, McPherson DC, Otte MM, Wang R, Ferguson CC, Eichenberger P, Driks A.** 2006. The *Bacillus subtilis* spore coat protein interaction network. Molecular microbiology **59:**487-502.

3. **Kuwana R, Kasahara Y, Fujibayashi M, Takamatsu H, Ogasawara N, Watabe K.** 2002. Proteomics characterization of novel spore proteins of *Bacillus subtilis*. Microbiology (Reading, England) **148:**3971-3982.

4. **Lai EM, Phadke ND, Kachman MT, Giorno R, Vazquez S, Vazquez JA, Maddock JR, Driks A.** 2003. Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*. Journal of bacteriology **185:**1443-1454.

5. **McKenney PT, Driks A, Eichenberger P.** 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol **11:**33-44.

6. **Henriques AO, T. V. Costa, L. O. Martins, and R. Zilhão.** 2004. Functional architecture and assembly of the spore coat, p. 65-85. *In* Ezio Ricca AOH, and Simon M. Cutting (ed.), Bacterial Spores: Probiotics and Emerging Applications. Horizon Scientific Press, London, UK.

7. **Henriques AO, Moran CP, Jr.** 2000. Structure and assembly of the bacterial endospore coat. Methods (San Diego, Calif **20:**95-110.

8. **Piggot PJ, Hilbert DW.** 2004. Sporulation of *Bacillus subtilis*. Current opinion in microbiology **7:**579-586.

9. **Higgins D, Dworkin J.** 2012. Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol Rev **36:**131-148.

10. **Driks A.** 1999. *Bacillus subtilis* spore coat. Microbiol Mol Biol Rev **63:**1-20.

11. **McKenney PT, Eichenberger P.** 2012. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. Molecular microbiology **83:**245-260.

12. **Takamatsu H, Kodama T, Nakayama T, Watabe K.** 1999. Characterization of the *yrbA* gene of *Bacillus subtilis*, involved in resistance and germination of spores. Journal of bacteriology **181:**4986-4994.

13. **Driks A, Roels S, Beall B, Moran CP, Jr., Losick R.** 1994. Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. Genes & development **8:**234-244.

14. **McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z, Eichenberger P.** 2010. A distance-weighted interaction map reveals a previously uncharacterized layer of the *Bacillus subtilis* spore coat. Curr Biol **20:**934-938.

15. **Zheng LB, Donovan WP, Fitz-James PC, Losick R.** 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. Genes & development **2:**1047-1054.

16. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2011. Proteins involved in formation of the outermost layer of *Bacillus subtilis* spores. Journal of bacteriology **193:**4075-4080.

17. **Wang KH, Isidro AL, Domingues L, Eskandarian HA, McKenney PT, Drew K, Grabowski P, Chua MH, Barry SN, Guan M, Bonneau R, Henriques AO, Eichenberger P.** 2009. The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*. Molecular microbiology 74:634-649.

18. **Seyler RW, Jr., Henriques AO, Ozin AJ, Moran CP, Jr.** 1997. Assembly and interactions of *cotJ*-encoded proteins, constituents of the inner layers of the *Bacillus subtilis* spore coat. Molecular microbiology **25:**955-966.

19. **Ishikawa S, Yamane K, Sekiguchi J.** 1998. Regulation and characterization of a newly deduced cell wall hydrolase gene (*cwlJ*) which affects germination of *Bacillus subtilis* spores. Journal of bacteriology **180:**1375-1380.

20. **Ragkousi K, Eichenberger P, van Ooij C, Setlow P.** 2003. Identification of a new gene essential for germination of *Bacillus subtilis* spores with Ca2+-dipicolinate. Journal of bacteriology **185:**2315-2329.

21. **Lorand L, Graham RM.** 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature reviews **4:**140-156.

22. **Zilhao R, Serrano M, Isticato R, Ricca E, Moran CP, Jr., Henriques AO.** 2004. Interactions among CotB, CotG, and CotH during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **186:**1110-1119.

23. **Geiser M, Cebe R, Drewello D, Schmitz R.** 2001. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. BioTechniques **31:**88-90, 92.

24. **Steinmetz M, Richter R.** 1994. Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in *Bacillus subtilis*, through *in vivo* recombination. Gene **142:**79-83.

25. **Zilhao R, Isticato R, Martins LO, Steil L, Volker U, Ricca E, Moran CP, Jr., Henriques AO.** 2005. Assembly and function of a spore coat-associated transglutaminase of *Bacillus subtilis*. Journal of bacteriology **187:**7753-7764.

26. **Miroux B, Walker JE.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. Journal of molecular biology **260:**289-298.

27. **Lemon KP, Grossman AD.** 2000. Movement of replicating DNA through a stationary replisome. Mol Cell **6:**1321-1330.

28. **Nicholson WL, Setlow P.** 1990. Sporulation, germination and outgrowth, p. 391-450. *In* Cutting CRHaSM (ed.), Molecular biology methods for *Bacillus*. John Wiley & Sons, Ltd., Chichester.

29. **Rasband WS** 1997-2012, posting date. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/. [Online.]

30. **Kuwana R, Okuda N, Takamatsu H, Watabe K.** 2006. Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. Journal of biochemistry **139:**887-901.

31. **Ragkousi K, Setlow P.** 2004. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. Journal of bacteriology **186:**5567-5575.

32. **Monroe A, Setlow P.** 2006. Localization of the transglutaminase cross-linking sites in the *Bacillus subtilis* spore coat protein GerQ. Journal of bacteriology **188:**7609-7616.

33. **Ozin AJ, Costa T, Henriques AO, Moran CP, Jr.** 2001. Alternative translation initiation produces a short form of a spore coat protein in *Bacillus subtilis*. Journal of bacteriology **183:**2032-2040.

34. **Takamatsu H, Kodama T, Imamura A, Asai K, Kobayashi K, Nakayama T, Ogasawara N, Watabe K.** 2000. The *Bacillus subtilis yabG* gene is transcribed by SigK RNA polymerase during sporulation, and *yabG* mutant spores have altered coat protein composition. Journal of bacteriology **182:**1883-1888.

35. **Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS biology **2:**e328.

36. **Steil L, Serrano M, Henriques AO, Volker U.** 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. Microbiology (Reading, England) **151:**399-420.

37. **Ozin AJ, Samford CS, Henriques AO, Moran CP, Jr.** 2001. SpoVID guides SafA to the spore coat in *Bacillus subtilis*. Journal of bacteriology **183:**3041-3049.

38. **Pinkas DM, Strop P, Brunger AT, Khosla C.** 2007. Transglutaminase 2 undergoes a large conformational change upon activation. PLoS biology **5:**e327.

39. **Stieler M, Weber J, Hils M, Kolb P, Heine A, Büchold C, Pasternack R, Klebe G.** 2013. Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. Angewandte Chemie International Edition **52:**11930-11934.

40. **Placido D, Fernandes CG, Isidro A, Carrondo MA, Henriques AO, Archer M.** 2008. Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. Protein expression and purification **59:**1-8.

41. **Tarcsa E, Candi E, Kartasova T, Idler WW, Marekov LN, Steinert PM.** 1998. Structural and transglutaminase substrate properties of the small proline-rich 2 family of cornified cell envelope proteins. The Journal of biological chemistry **273:**23297-23303.

42. **Aeschlimann D, Paulsson M.** 1991. Cross-linking of laminin-nidogen complexes by tissue transglutaminase. A novel mechanism for basement membrane stabilization. The Journal of biological chemistry **266:**15308-15317.

43. **Candi E, Tarcsa E, Idler WW, Kartasova T, Marekov LN, Steinert PM.** 1999. Transglutaminase cross-linking properties of the small proline-rich 1 family of cornified cell envelope proteins. Integration with loricrin. The Journal of biological chemistry **274:**7226-7237.

44. **Barry EL, Mosher DF.** 1990. Binding and degradation of blood coagulation factor XIII by cultured fibroblasts. The Journal of biological chemistry **265:**9302-9307.

45. **Birckbichler PJ, Orr GR, Carter HA, Patterson MK, Jr.** 1977. Catalytic formation of epsilon-(gamma-glutamyl)lysine in guinea pig liver transglutaminase. Biochemical and biophysical research communications **78:**1-7.

46. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2010. Localization of proteins to different layers and regions of *Bacillus subtilis* spore coats. Journal of bacteriology **192:**518-524.

47. **Takamatsu H, Imamura D, Kuwana R, Watabe K.** 2009. Expression of *yeeK* during *Bacillus subtilis* sporulation and localization of YeeK to the inner spore coat using fluorescence microscopy. Journal of bacteriology **191:**1220-1229.

48. **Sauvage E, Kerff Fdr, Terrak M, Ayala JA, Charlier P.** 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiology Reviews **32:**234-258.

49. **Scheffers DJ, Pinho MG.** 2005. Bacterial Cell Wall Synthesis: New Insights from Localization Studies. Microbiology and Molecular Biology Reviews **69:**585-607.

50. **Pinho MG, Errington J.** 2005. Recruitment of penicillin-binding protein PBP2 to the division site of *Staphylococcus aureus* is dependent on its transpeptidation substrates. Molecular microbiology **55:**799-807.

51. **Potluri L, Karczmarek A, Verheul J, Piette A, Wilkin J-M, Werth N, Banzhaf M, Vollmer W, Young KD, Nguyen-Distèche M, Den Blaauwen T.** 2010. Septal and lateral wall localization of PBP5, the major D,D-carboxypeptidase of *Escherichia coli*, requires substrate recognition and membrane attachment. Molecular microbiology 77:300-323.

52. **Costa T, Priyadarshini R, Jacobs-Wagner C.** 2008. Localization of PBP3 in *Caulobacter crescentus* is highly dynamic and largely relies on its functional transpeptidase domain. Molecular microbiology **70:**634-651.

53. **Lages MCA, Beilharz K, Morales Angeles D, Veening J-W, Scheffers D-J.** 2013. The localization of key *Bacillus subtilis* penicillin binding proteins during cell growth is determined by substrate availability. Environmental microbiology **15:**3272-3281.

54. **Pinho MG, Kjos M, Veening J-W.** 2013. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. Nature Reviews Microbiology **11:**601-614.

55. **Kobayashi K, Hashiguchi K, Yokozeki K, Yamanaka S.** 1998. Molecular cloning of the transglutaminase gene from *Bacillus subtilis* and its expression in *Escherichia coli*. Bioscience, biotechnology, and biochemistry **62:**1109-1114.

56. **Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE.** 1994. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Sci **3:**1131-1135.

57. **Lorand L.** 2001. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. Ann N Y Acad Sci **936:**291-311.

Chapter 5

General discussion

The coat is, in many sporeformers, the outermost layer of the spore. It is one of the most robust protein assemblies known, and is, not only, a key determinant in the resistance properties of the spore, but also in its functional interactions with the immediate environment. While great stride has been made to understand the assembly process of the coat, the focus has been put on the role of the morphogenetic proteins while the role of individual non-morphogenetic proteins is largely unknown. Despite the widespread distribution of transglutaminases (TGases) and their important contribution in development processes, studies on bacterial TGases are scarce. While one such enzyme, Tgl, had been identified as a factor in proper spore coat formation, studies on its role and assembly in spore morphogenesis and on its structural and mechanistic characteristics were lacking. Thus, we sought to contribute for the knowledge of both, spore coat formation and bacterial TGases, by analyzing Tgl.

First, we have seen that Tgl represents a striped down TGase, and can be considered as the current representative of the minimal transglutaminase (Chapter 2). Also, Tgl appears to catalyze protein cross-linking through a Cys-Glu catalytic diad. Second, our work has provided evidence that Tgl is present at the inner coat and cortex layers of the spore (Chapter 3). Its subcellular localization at the cortex, along with that of SafA and C30, appears to be modulated by an interdependent auto-regulatory loop. The fact that Tgl acts on cortex-associated proteins which are also present in the coat suggests, that the enzyme acts to cement the cortex/inner coat interface. Thus, Tgl cross-links a pre-assembled structure to confer extra stability to the overall spore. Finally, we have uncovered a new assembly mechanism not previously recognized in coat formation, where an enzyme is recruited to a forming structure via its substrates (Chapter 4). While the current view on spore coat formation emphasizes the role of morphogenetic proteins and the different waves of the encasement process, evidence suggests that several other factors are at play. The fact that Tgl is recruited to the spore by, at least, some of its substrates is a prime example of an assembly mechanism operating downstream of the morphogenetic effectors. Strong redundancy in the assembly of the Tgl, conferred by its various substrates, likely reflects the need for Tgl localization in the spore.

Like Tgl, it is possible that the recruitment of other proteins to the forespore will depend on yet unidentified assembly mechanisms.

TGL AS THE MINIMAL TGASE: IMPLICATIONS FOR TGL FUNCTION

The minimal features for the catalysis of protein cross-linking reactions. Due to their protein cross-linking role, the activity of TGases is highly regulated. Animal TGases need calcium as co-factor, and are frequently synthesized in zymogenic form, or inhibited by GTP/GDP or ATP binding (1-3). Even the "simpler" MTG, a microbial TGase, is synthesized as a zymogen, secreted to the outside of the cell, and, only then, activated by processing (4-6). However, previous *in vitro* work had already suggested that Tgl is produced in active form (7). This inference was confirmed by the unraveling of the Tgl structure (Chapter 2), which clearly displays the catalytic Cys accessible from the outside environment. The synthesis of Tgl in active form is further supported by the localization of the active site in the middle of an hydrophobic tunnel that traverses the molecule from side to side (Figure 2.4). While the catalytic Cys is inaccessible in the inactive forms of animal TGases, a tunnel that leads to the catalytic Cys is detected in the active forms of TGase 2 and 3, and Factor XIIIa (FXIIIa), and is likely a common feature of the activated forms of animal-type TGases (8-10). Because no structural changes seem necessary to activate Tgl, this bacterial enzyme can be viewed as a stripped down TGase endowed with what appears to be the minimal features for the catalysis of protein cross-linking. In line with this view, while animal TGases are multi-domain enzymes (Figure 1.5), the "extra", non-catalytic, domains appear to be important for proper localization, or to exert a regulatory role, e.g., to maintain the enzymes in an inactive state. For example, the N-terminal domain of TGase 1 corresponds to a membrane anchoring domain (3,10), while the N-terminal domain of TGase 2 is important for its extracellular localization at the cell surface (11); also, the C-terminal of TGase 2 contains the residues involved in GTP/GDP binding (12) (Figure 1.5Ab). In accordance, a study with TGase 1 (a 816 residue protein) showed that, a mutant enzyme from which, the first 109 and the last 240 residues where eliminated, and thus containing mainly its

conserved catalytic core, still retains enzymatic activity comparable to the wild type enzyme (13).

Activity regulation is coupled to substrate specificity. Because Tgl is synthesized in active form, a major question that arises is how its activity is regulated. Several of our results point to the view that Tgl has a strict substrate specificity. First, of the more than 70 polypeptides that compose the spore coat, Tgl only modifies a small number. Second, unlike other TGases (14,15), Tgl can be produced to high yields in active form in *E. coli* without causing cellular lysis (Chapters 2 and 4). Finally, Tgl's ability to label exogenous proteins, such as BSA, is not as proficient as other TGases. We base this last inference on the experimental conditions used to characterize the activity of Tgl by dansylcadaverine labeling (Chapters 2 and 4), in which the amount of Tgl had to be increased by 10 fold relative to the concentration used for other TGases (16,17) for the detection of labeled BSA (Figures 2.7A and 4.4). Moreover, the excess of dansylcadaverine inhibited BSA cross-linking (not shown). In sharp contrast, under similar conditions and using two of Tgl's physiological substrates, SafA or C30, cross-linking was not inhibited and both cross-linking and labeling reactions were simultaneous detected (not shown).

This seemingly strict specificity of Tgl is most likely imparted by residues located at both entrances of the tunnel. Because the catalytic residues of Tgl (defined in Chapter 2) are hidden within the tunnel of the enzyme, both entrances become ports of entry for the substrates. Also, the tunnel is quite narrow and the two entrances are on opposite sides, thus the tunnel physically separates the docking sides of Gln and Lys donor substrates (Q and K substrates, respectively). Hence, it seems likely that residues located at both entrances of the tunnel have a role in substrate interactions/specificity. Accordingly, the exchange of residues at both entrances of the tunnel to Ala severally impairs enzymatic activity (Figures 2.7, 4.4 and 4.8A). In agreement with the role of these residues in substrate recognition and not directly in catalysis, Tgl^{W149A} still retained considerable activity but its specificity was altered (Figures 4.5 and 4.8B). Likewise, in striking contrast to the wild type enzyme, the expression of Tgl^{F69A} in *E. coli* BL21(DE3) caused

cellular lysis (not shown), which suggests that the mutant enzyme is atypically recognizing *E. coli* proteins as substrates.

Impact on the localization mechanism of Tgl. Interestingly, the high specificity of Tgl also appears to play a role in the localization of the enzyme during sporulation. Our results support the view that the recruitment of Tgl to the forespore is dependent on its substrates and on residues at the entrances of the Tgl tunnel (Chapter 4). Furthermore, front side residues have a pivotal role in localization, and our results indicate that the front side entrance of the tunnel corresponds to the docking site of the Q substrate (Chapter 4). However, catalytic, and back entrance residues, also play a part in the assembly of the enzyme. Thus we propose that the interaction of Tgl with a Q substrate is a key step in assembly; reaction of the Q substrate with the catalytic Cys to form an acyl-enzyme intermediate also plays a role, perhaps by stabilizing the complex (steps a to c of the reaction mechanism of TGases, Figure 2.1). A second interaction is then made at the back entrance of the tunnel, with either a different region of the bound Q substrate, acting as a Lys donor, or with a different K substrate.

Activity regulation by substrate alterations. The realization that the formation of the acyl-enzyme intermediate is likely to be important for the recruitment of Tgl to the forespore was surprising, as the activity of the enzyme is only detected very late in sporulation, possibly dependent on coat maturation processes that occur after mother cell lysis (18,19). Thus the formation of the ε-(γ-glutamyl)lysyl isopeptide bond has to be somehow delayed. One possibility is that Tgl, like FXIIIa, recognizes its substrates in a specific structural conformation, which in the case of the FXIIIa substrates is triggered by their processing (1,20) (see also Chapter 1). Interestingly, a functional relationship between Tgl and the YabG protease has been proposed (21,22). In the absence of YabG, the known substrates of Tgl are more extractable and, apart from GerQ, they appear to be present in an unprocessed form (21). Additionally, in spores of a *yabG* mutant, Tgl is present in similar levels as compared to wild type spores (not shown). Thus, one possibility is that YabG processes the substrates of Tgl which can then be used in the protein cross-linking reaction. However, we have identified

the processing site of both YeeK and C30, and our results show that processing is not a strict requirement for Tgl-dependent cross-linking (not shown). Thus we propose that late maturation coat processes, not directly resulting from processing, and to which YabG likely contributes, will induce conformational changes in the substrates of Tgl, so that Tgl-dependent cross-linking is triggered.

HOW DOES TGL REACH THE CORTEX?

In Chapter 3 we have seen that Tgl localizes mainly to the cortex and to cortex/inner coat interface (Figure 3.5), and that its cortex localization is dependent on *safA*. SafA is also detected at the cortex/inner coat interface (23), and SafA, C30, and Tgl are stably associated with the cortex layer in wild type spores (Figures 3.3 and 3.4A). One major unsolved question is how these proteins reach the cortex region. Following engulfment completion the forespore is surrounded by a double membrane system (Figures 1.1C and 1.3) and the cortex will be assembled in the intermembrane space. The inner forespore membrane is kept in free spores and it will become the cytoplasmic membrane of the new cell generated by spore germination. The fate of the outer forespore membrane is not known (24,25). Thus, either SafA/C30, and Tgl are transported across the outer forespore membrane to the cortex, or this membrane is absent by the time the proteins localize to the cortex. Since Tgl shows no obvious secretion signals (or any of the characteristics of secreted proteins; below), we presently favor a scenario in which Tgl reaches the cortex at late stages in sporulation, when the outer forespore membrane is supposedly lost.

The same might be true for SafA and C30. However, an alternative view is that SafA crosses the outer forespore membrane and is tethered to the cortex by its LysM domain (peptidoglycan-binding domain (26)). Although SafA does not have a signal peptide, there is increasing evidence that proteins without a typical signal peptide can be actively secreted across the membrane by unknown mechanisms (27-30). Studies have also shown that secreted and cytoplasmic proteins have distinct intrinsic characteristics (such as their amino acid composition or predicted disordered regions)

(27,30), which has led to the creation of a prediction method denoted SecretomeP (27). Using the SecretomeP 2.0 server, SafA, but not Tgl, is predicted to be a secreted protein.

Perhaps the most striking evidence in support of the view that SafA may be secreted across the outer forespore membrane comes from our initial experiments, comparing the original *safA* mutant (Δ*safA*$_{spc}$) with the newly constructed in-frame deletion (Δ*safA*$_{46\text{-}249}$) mutant (Figure 3.1). While the in-frame *safA* deletion mutant could be complemented to restore the wild type phenotype, the same was not true for the *safA*$_{spc}$ mutant (Figure 3.1A). We analyzed the levels of SafA in the coat and lysozyme fractions of purified spores of both complemented strains (Figure 5.1). Our results show that, in the Δ*safA*$_{spc}$ *safA* strain, SafA/C30 localization in the cortex layer is lost, while the same is not observed for the complemented in-frame deletion mutant (Δ*safA*$_{46\text{-}249}$ *safA*, Figure 5.1). This indicates that the defects in the SafA/C30-cortex association detected in strain Δ*safA*$_{spc}$ *safA* (Figure 5.1) are probably due to polar effects.

The *coxA* gene is located downstream of *safA*. It has been previously demonstrated that *coxA* expression is under the control of two promoters, and, although *coxA* is transcribed alone (σ^G dependent), it is also cotranscribed with *safA* under the control of the σ^E promoter of *safA* (31,32). In addition, the interruption of *safA* by a single cross-over event leads to the expression of *coxA* from only its σ^G promoter (31). Thus, it seems possible that the expression of *coxA* from both the σ^E and σ^G promoters is important for the proper SafA-cortex association. The C-terminal domain of CoxA is predicted (33) to share structural homology with the C-terminal domain of EscJ from enteropathogenic *E. coli* (with a confidence of 88.4 %; pdb: 1YJ7). EscJ belongs to the YscJ/PrgK family of proteins involved in the formation of the needle complex in type 3 secretion systems. EscJ most likely oligomerizes to form a 24-subunit ring that lies on top of the inner membrane of *E. coli*, recruiting other partners to form the final secretion apparatus (34). Considering the homology detected between CoxA and EscJ, it is possible that CoxA will also be able to form a multimer. Furthermore, another sporulation protein, SpoIIIAH, was also initially identified as an EscJ homologue based on structural predictions (35). Later,

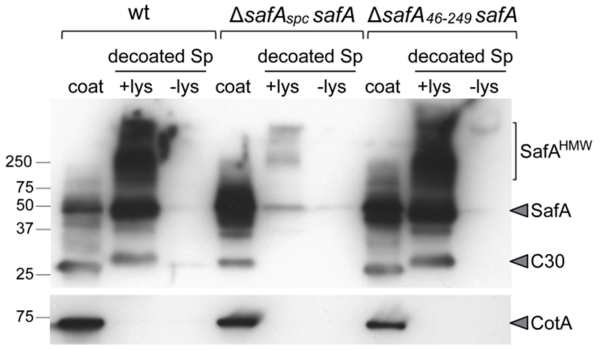


Figure 5.1. The complementation of the *safA*$_{spc}$ mutant does not restore the cortex association of SafA and C30. Two *safA* mutant strains were constructed (for details, see Chapter 3). The first had the *safA* gene interrupted by the spectinomycin resistance cassette (Δ*safA*$_{spc}$), the second corresponded to an in-frame deletion of codons 46 to 249 of *safA* (Δ*safA*$_{46-249}$). Δ*safA*$_{spc}$ was complemented with *safA*, by a single cross-over event, at the *safA* locus (Δ*safA*$_{spc}$ *safA*), while strain Δ*safA*$_{46-249}$ was complemented by the introduction of *safA* at the *amyE* locus (under the control of the native *safA* promoter, Δ*safA*$_{46-249}$ *safA*). Both complemented strains along with the wild type, MB24, were grown in DSM until 24 hr after sporulation initiation. The spores were purified, and the levels of SafA/C30 in the coat and cortex fractions analyzed by immunoblot analysis with an anti-SafA antibody (as described in Figure 3.3). The image clearly shows that, in the Δ*safA*$_{spc}$ *safA* strain the fraction of SafA/C30 usually associated with the cortex (+lys lane) is mostly absent, while the same does not happen in the Δ*safA*$_{46-249}$ *safA* strain. SafA^{HMW} refers to high molecular weight products detected with anti-SafA. The site of migration of the molecular weight markers (kDa) is shown on the left hand side.

the structure of SpoIIIAH was solved and the homology to EscJ confirmed (36,37). SpoIIIAH (whose transcription is dependent on σ^E) and SpoIIQ (under the control of σ^F), multimerize into a ring to form the core elements of a channel that spans across the inner and outer forespore membranes during engulfment (35,38,39). Like SpoIIIAH and EscJ, CoxA may also create a scaffold for the recruitment of other proteins to create a channel across the intermembrane space. Since SafA recruits Tgl to the cortex, an intriguing possibility is that CoxA may also be required for the assembly of Tgl. Whether or not CoxA is involved in the transport of SafA, C30, and Tgl to the cortex region remains to be tested.

TGL CEMENTS THE CORTEX/INNER COAT INTERFACE

Tgl, SafA, C30, and GerQ have been shown, or postulated, to be present at the cortex/inner coat interface (Chapter 3 and (23,24)). In *B. anthracis*, a substrate of Tgl is a protein that localizes to the cortex (40), and preliminary evidence suggests that YeeK localizes, either to the cortex, or in close proximity to it (not shown), in accordance with previous results (41). The subcellular localization of Tgl and its substrates led us to speculate that the

enzyme has a role in cementing the interface between the cortex and inner coat layers. Additionally, it has been shown that proper cortex and coat formation are coupled (42). If Tgl is important for the stable and enduring connection between the inner coat and cortex, the weakening of this link could have repercussions on the juxtaposed outer coat layer. Accordingly, in *tgl* mutant spores, the outer coat layer is affected (18), a layer where Tgl seems to be present at only low levels (Figure 3.5). Also, the absence of Tgl decreases the wet heat resistance of spores similar to what is seen in double *cotE gerE* spores (which lack a coat and in which only a small rind of material is detected around the spore) (43,44).

BACTERIAL TRANSGLUTAMINASES AND TGL

The current knowledge on bacterial TGases is very scarce and few reports exist on this class of enzymes. The only studies conducted on bacterial enzymes which either present TGase activity or a TGase-like fold, are for: MTG, from *Streptomyces mobaraensis* (or MTG homologues of different *Streptomyces* species); two toxins, the cytotoxic necrotizing factor 1 (CNF1), from *Escherichia coli* and the *Bordetella* dermonecrotic toxin (DNT); WbmE, from *Bordetella bronchiseptica;* and TgpA, from *Pseudomonas aeruginosa*.

MTG. MTG has been the most extensively researched bacterial TGase and is also the only current representative of a bacterial TGase in the PDB database. Yet, the studies on MTG have focused on an applied perspective and the biological role of MTG is unknown. It is however believed that the enzyme is involved in formation of the aerial mycelium in *S. mobaraensis*, which is the site of spore formation (4,5,45). In a clear similarity to Tgl, MTG is also a single domain, calcium-independent enzyme. Nevertheless, the overall structure of both enzymes is very different (Figure 5.2A,B). MTG presents a disk like shape with its active site located on a deep cleft, and residues within this cleft appear to be important for substrate recognition (Figure 5.2B) (46,47). The wide cleft in MTG warrants a much broader specificity for the Q substrate when compared to animal TGases (48,49), which once again strengthens the idea that the tunnel of Tgl confers a stringent specificity to the enzyme.

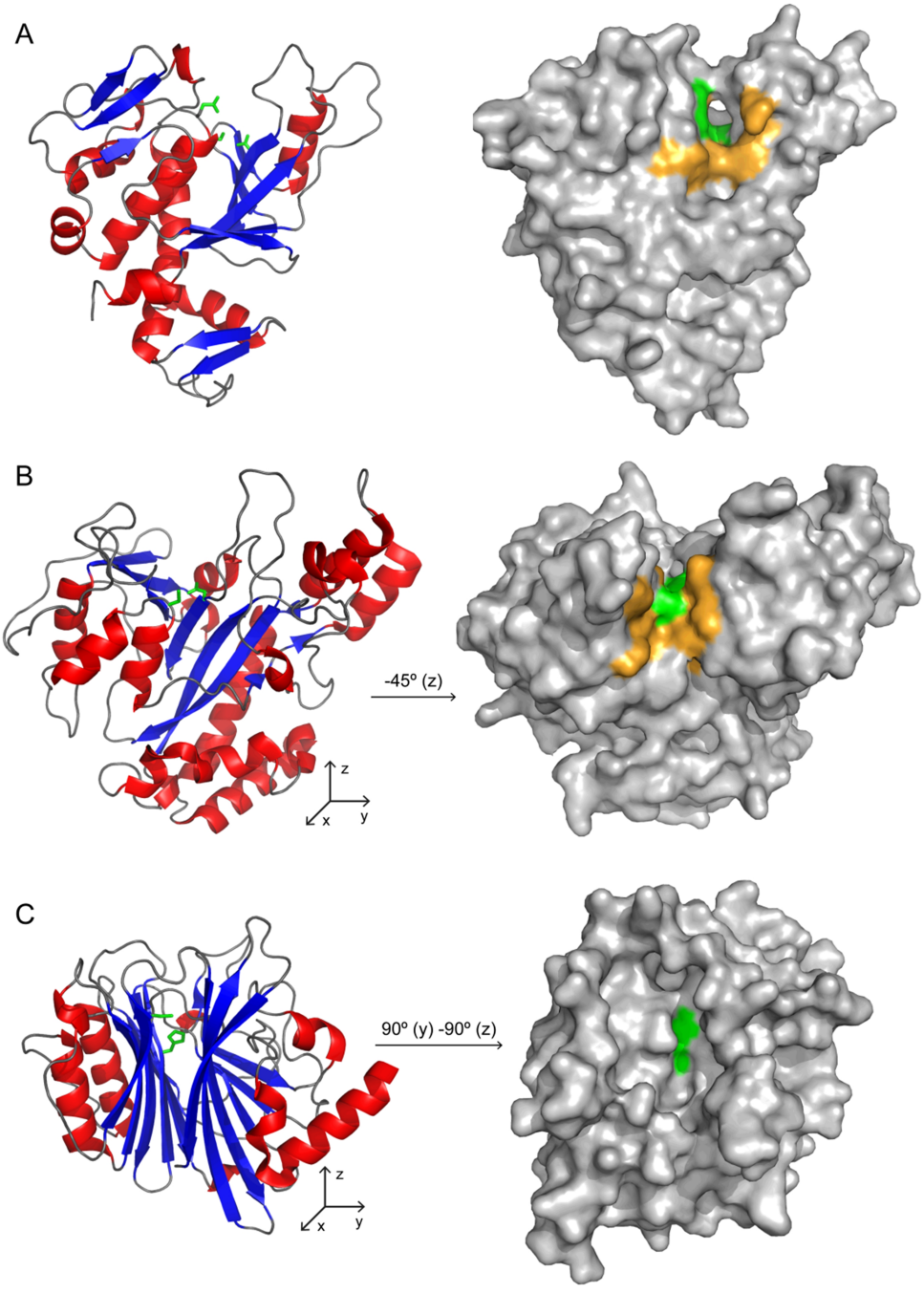


Figure 5.2. Structural representation of the front side of Tgl (A), mature MTG (B), and C-terminal catalytic domain of CNF1 (C). The secondary structure elements are represented on the left hand side of each panel, while the surface representation is shown on the right-hand side. Active site residues are shown in green, and the side chains are displayed as sticks on the left hand side images. Residues most likely involved in substrate interactions are shown in orange. In B and C, the structures on the right hand side have been rotated according to the axes and relative to the image on the left hand side. The structure of Tgl corresponds to the one presented in Chapter 2; MTG and CNF1 correspond to pdb files 1IU4 and 1HQ0, respectively.

CNF1 and DNT. These two toxins target the eukaryotic Rho proteins, which are GTPases involved in the regulation of the actin cytoskeleton and in a wide range of signaling events. CNF1 and DNT, deaminate a specific Gln residue of the Rho proteins which is necessary for catalysis of GTP hydrolysis. This renders the Rho proteins constitutively active, which leads to cytotoxic effects (50). The two proteins share a region of homology located at the C-terminal which corresponds to the catalytic domain, and is sufficient to cause the cytotoxic effects seen with CNF1 and DNT (50-52). It has also been shown that both toxins have *in vitro* transamidation activity for the same residue which is deaminated *in vivo* (51-53). In fact, DNT appears to favor transamidation over deamination (51,52). The *in vivo* transamidation activity of DNT has also been reported, which causes the inhibition of GTP hydrolysis by RhoA, just like RhoA deamination (53).

The structure of the C-terminal domain of CNF1 has been solved (Figure 5.2C), which is likely identical to that of DNT, and it appears that these toxins (and homologues) constitute a family with a unique fold (54). The unique fold of CNF1 is most likely related to the strict specificity of these toxins. While human TGases can modify several of the Gln residues of Rho proteins, CNF1 and DNT only recognize one, and the two toxins are unable to use casein, actin or fibronectin as substrates, which can be typically used by TGases (51,52). The catalytic residues of CNF1 sit at the bottom of a very deep pocket whose dimensions exclude arbitrary substrates (Figure 5.2C). Furthermore, it has been postulated that catalysis only occurs after a conformation change which is derived by an induced fit between Rho and CNF1 (54). Interestingly, both CNF1 and DNT, are calcium-independent enzymes and the deamination/transamidation reactions are catalyzed by a Cys-His catalytic diad (52). While the overall fold of TGases and CNF1 are very different, the Cys and His residues of both groups of enzymes can be superimposed (54). The differences in activity between, CNF1/DNT and known TGases, is most likely a reflection of their contrasting structures. *In vivo*, CNF1 is an deamidase and while DNT has *in vivo* transamidation activity, the enzyme does not catalyze protein cross-linking reactions and the second substrates *in vivo* corresponds to L-lysine or small polyamines (spermidine or putrescine) (53). Thus, while the pocket in CNF1/DNT can

accommodate Rho proteins (Q substrate) it may be difficult to allow the binding of second protein (K substrate).

WbmE and TgpA. These two enzymes were identified in a PSI-BLAST study as possible TGases (55). This study identified homologues of animal TGases in archaea, bacteria and eukaryotes based on motif conservation around the catalytic residues corresponding to those of the catalytic triad of FXIIIa (55). In line with the unique sequence of Tgl, the enzyme was not detected in the study. Although the newly identified proteins are thought to compose a superfamily of protein homologues of human TGases, their role as TGases remains to be tested. Furthermore, some of the proteins identified had already been shown to be peptidases and it is likely that many of the pinpointed members of this superfamily will be proteases (55,56).

Recently, two studies have characterized: WbmE and TgpA (57,58). Both proteins appear to function at the periplasm, and TgpA remains membrane bound. WbmE functions as a deaminase of the lipopolysaccharide present in the outer membrane (57). TGases are known to catalyze deamination reactions *in vivo* (2). However, considering the two very different substrates of TGases and WbmE (proteins vs uronamide sugars), it is unlikely that WbmE has TGase activity. Thus, the catalytic core of WbmE is predicted by the Phyre2 server (33), to have structural homology to a putative cysteine protease, an N-acetyltransferase or a peptide N-glycanase (best three hits with about 94, 77 and 67% confidence, corresponding to pdb files 3ISR, 2PFR and 2F4M, respectively).

On the other hand, the proposed TGase domain of TgpA is predicted to have a strong homology with animal TGases (33). Of the first 10 hits (all with more than 98% confidence), 8 correspond to the catalytic cores found in the crystal structures of TGase 2, FXIII, FTG (fish-derived TGase), and TGase 3. In accordance to the high homology to animal TGases, it was confirmed that TgpA has *in vitro* TGase activity, which is calcium-independent. The physiological role of TgpA is not known but the enzyme is necessary for viability of *P. aeruginosa*, suggesting its involvement in cell wall metabolism (58). Despite the strong predicted structural homology of TgpA with animal TGases, in particular with the structure of activated TGase 2, the TgpA

structure generated by Phyre2 does not reveal a tunnel that leads to the catalytic Cys. Rather, this residue is predicted to be solvent exposed. However, the TGase core of TgpA (the sequence of the protein with homology to TGases) corresponds to a stretch of ~70 residues flanked by two other sections of 70 to 80 residues for which no structural homology can be detected. It is possible that these flanking regions shield the catalytic Cys (whether forming a tunnel or not) controlling in that manner the activity of the enzyme.

A global view of bacterial TGases (or TGase-like proteins). As stated above, all of the enzymes previously described seem to function via a catalytic diad (with the exception of WbmE and TgpA for which this is not known) and their activity appears to be calcium independent (this has not been tested for WbmE). Thus, these two features, in a striking contrast to animal TGases, may prove to be the norm in bacterial TGases or related enzymes. Calcium binding in animal TGases induces conformational changes in the active site of the enzymes (Figure 1.5), so that the catalytic Cys becomes accessible from the outside environment (10,20,59). Tgl, CNF1, and DNT are synthesized in active form. While MTG is synthesized as a zymogen, the removal of the pro-sequence (which forms an α-helix that covers the active site cleft) does not lead to structural alterations on the active site core (48). Thus, once bacterial TGases are "activated" their activity has to be regulated by mechanisms other than co-factor binding. In Tgl, CNF1 and DNT, this is probably ensured by the strict specificity of the enzymes, conveyed by the tunnel of Tgl, or the deep pocket seen in the toxins (Figure 5.2A,C). While the same strategy is adopted by Tgl and the toxins, their structure is very different, likely reflecting the different K substrates used by the enzymes (proteins vs small amines, see above). The activity of Tgl also seems to be controlled by alterations on its substrates and this has also been reported for MTG (at least for one of its possible physiological substrates) (45).

Finally, apart from Tgl, all the enzymes mentioned above are secreted. This non-cytoplasmic localization is, in itself, a way to regulate their enzymatic activity. On the contrary, Tgl is a cytoplasmic protein and it is also

the only bacterial enzyme where a tunnel is seen. Thus, it is tempting to propose that the fold of Tgl is adopted by bacterial cytoplasmic TGases in which the tunnel will regulate not only, the activity of the enzyme, but also its localization within the cell. Nonetheless, such a possibility can only be assessed when further studies on bacterial TGases become available.

REFERENCES

1. **Lorand L, Graham RM.** 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature reviews **4:**140-156.

2. **Gundemir S, Colak G, Tucholski J, Johnson GVW.** 2012. Transglutaminase 2: A molecular Swiss army knife. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1823:**406-419.

3. **Klöck C, Khosla C.** 2012. Regulation of the activities of the mammalian transglutaminase family of enzymes. Protein Science **21:**1781-1791.

4. **Zotzel J, Keller P, Fuchsbauer HL.** 2003. Transglutaminase from *Streptomyces mobaraensis* is activated by an endogenous metalloprotease. European journal of biochemistry / FEBS **270:**3214-3222.

5. **Zotzel J, Pasternack R, Pelzer C, Ziegert D, Mainusch M, Fuchsbauer HL.** 2003. Activated transglutaminase from *Streptomyces mobaraensis* is processed by a tripeptidyl aminopeptidase in the final step. European journal of biochemistry / FEBS **270:**4149-4155.

6. **Pasternack R, Dorsch S, Otterbach JT, Robenek IR, Wolf S, Fuchsbauer HL.** 1998. Bacterial pro-transglutaminase from *Streptoverticillium mobaraense*-purification, characterisation and sequence of the zymogen. European journal of biochemistry / FEBS **257:**570-576.

7. **Placido D, Fernandes CG, Isidro A, Carrondo MA, Henriques AO, Archer M.** 2008. Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. Protein expression and purification **59:**1-8.

8. **Pinkas DM, Strop P, Brunger AT, Khosla C.** 2007. Transglutaminase 2 undergoes a large conformational change upon activation. PLoS biology **5:**e327.

9. **Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM.** 2002. Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. The EMBO journal **21:**2055-2067.

10. **Boeshans KM, Mueser TC, Ahvazi B.** 2007. A three-dimensional model of the human transglutaminase 1: insights into the understanding of lamellar ichthyosis. Journal of molecular modeling **13:**233-246.

11. **Gaudry CA, Verderio E, Aeschlimann D, Cox A, Smith C, Griffin M.** 1999. Cell surface localization of tissue transglutaminase is dependent on a fibronectin-binding site in its N-terminal beta-sandwich domain. The Journal of biological chemistry **274:**30707-30714.

12. **Liu S, Cerione RA, Clardy J.** 2002. Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. Proceedings of the National Academy of Sciences of the United States of America **99:**2743-2747.

13. **Kim SY, Kim IG, Chung SI, Steinert PM.** 1994. The structure of the transglutaminase 1 enzyme. Deletion cloning reveals domains that regulate its specific activity and substrate specificity. The Journal of biological chemistry **269:**27979-27986.

14. **Reiss K, Kirchner E, Gijzen M, Zocher G, Loffelhardt B, Nurnberger T, Stehle T, Brunner F.** 2011. Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria. Journal of Biological Chemistry **286:**42585-42593.

15. **Yokoyama KI, Nakamura N, Seguro K, Kubota K.** 2000. Overproduction of microbial transglutaminase in *Escherichia coli*, *in vitro* refolding, and characterization of the refolded form. Bioscience, biotechnology, and biochemistry **64:**1263-1270.

16. **Sugimura Y, Hosono M, Wada F, Yoshimura T, Maki M, Hitomi K.** 2006. Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGase 2 and Factor XIIIa. The Journal of biological chemistry **281:**17699-17706.

17. **Sugimura Y, Yokoyama K, Nio N, Maki M, Hitomi K.** 2008. Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. Archives of biochemistry and biophysics **477:**379-383.

18. **Zilhao R, Isticato R, Martins LO, Steil L, Volker U, Ricca E, Moran CP, Jr., Henriques AO.** 2005. Assembly and function of a spore coat-associated transglutaminase of *Bacillus subtilis*. Journal of bacteriology **187:**7753-7764.

19. **Ragkousi K, Setlow P.** 2004. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. Journal of bacteriology **186:**5567-5575.

20. **Lorand L.** 2001. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. Ann N Y Acad Sci **936:**291-311.

21. **Kuwana R, Okuda N, Takamatsu H, Watabe K.** 2006. Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. Journal of biochemistry **139:**887-901.

22. **Takamatsu H, Imamura A, Kodama T, Asai K, Ogasawara N, Watabe K.** 2000. The *yabG* gene of *Bacillus subtilis* encodes a sporulation specific protease which is involved in the processing of several spore coat proteins. FEMS microbiology letters **192:**33-38.

23. **Ozin AJ, Henriques AO, Yi H, Moran CP, Jr.** 2000. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. Journal of bacteriology **182:**1828-1833.

24. **Setlow P.** 2014. Germination of Spores of *Bacillus* Species: What We Know and Do Not Know. Journal of bacteriology **196:**1297-1305.

25. **Leggett MJ, McDonnell G, Denyer SP, Setlow P, Maillard JY.** 2012. Bacterial spore structures and their protective role in biocide resistance. Journal of applied microbiology **113:**485-498.

26. **Buist G, Steen A, Kok J, Kuipers OP.** 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. Molecular microbiology **68:**838-847.

27. **Bendtsen JD, Kiemer L, Fausboll A, Brunak S.** 2005. Non-classical protein secretion in bacteria. BMC Microbiol **5:**58.

28. **Holland IB.** 2010. The extraordinary diversity of bacterial protein secretion mechanisms, p. 1-20. *In* Economou A (ed.), Protein Secretion, Methods and Protocols, 2010/04/27 ed, vol. 619. Humana Press.

29. **Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E, Dorenbos R, Dubois JY, Westers H, Zanen G, Quax WJ, Kuipers OP, Bron S, Hecker M, van Dijl JM.** 2004. Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev 68:207-233.

30. **Yang CK, Ewis HE, Zhang X, Lu CD, Hu HJ, Pan Y, Abdelal AT, Tai PC.** 2011. Nonclassical Protein Secretion by *Bacillus subtilis* in the Stationary Phase Is Not Due to Cell Lysis. Journal of bacteriology **193:**5607-5615.

31. **Takamatsu H, Kodama T, Nakayama T, Watabe K.** 1999. Characterization of the *yrbA* gene of *Bacillus subtilis*, involved in resistance and germination of spores. Journal of bacteriology **181:**4986-4994.

32. **Eichenberger P, Jensen ST, Conlon EM, van Ooij C, Silvaggi J, Gonzalez-Pastor JE, Fujita M, Ben-Yehuda S, Stragier P, Liu JS, Losick R. 2003.** The sigmaE regulon and the identification of additional sporulation genes in *Bacillus subtilis*. Journal of molecular biology 327:945-972.

33. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc **4:**363-371.

34. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, Pfuetzner RA, Frey EA, Brett Finlay B, Miller SI, Strynadka NCJ. 2005. Structural characterization of the molecular platform for type III secretion system assembly. Nature 435:702-707.

35. **Meisner J, Wang X, Serrano M, Henriques AO, Moran CP, Jr.** 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. Proceedings of the National Academy of Sciences of the United States of America **105:**15100-15105.

36. **Levdikov VM, Blagova EV, McFeat A, Fogg MJ, Wilson KS, Wilkinson AJ.** 2012. Structure of components of an intercellular channel complex in sporulating *Bacillus subtilis*. Proceedings of the National Academy of Sciences **109:**5441-5445.

37. **Meisner J, Maehigashi T, Andre I, Dunham CM, Moran CP, Jr.** 2012. Structure of the basal components of a bacterial transporter. Proceedings of the National Academy of Sciences of the United States of America **109:**5446-5451.

38. **Camp AH, Losick R.** 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*. Genes & development **23:**1014-1024.

39. **Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Jr., Rudner DZ.** 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. PLoS Genet **5:**e1000566.

40. **Moody KL, Driks A, Rother GL, Cote CK, Brueggemann EE, Hines HB, Friedlander AM, Bozue J.** 2009. Processing, assembly and localization of a *Bacillus anthracis* spore protein. Microbiology (Reading, England) **156:**174-183.

41. **Imamura D, Kuwana R, Takamatsu H, Watabe K.** 2010. Localization of proteins to different layers and regions of *Bacillus subtilis* spore coats. Journal of bacteriology **192:**518-524.

42. **Ebmeier SE, Tan IS, Clapham KR, Ramamurthi KS.** 2012. Small proteins link coat and cortex assembly during sporulation in *Bacillus subtilis*. Molecular microbiology **84:**682-696.

43. **Sanchez-Salas JL, Setlow B, Zhang P, Li Yq, Setlow P.** 2011. Maturation of Released Spores Is Necessary for Acquisition of Full Spore Heat Resistance during *Bacillus subtilis* Sporulation. Applied and environmental microbiology **77:**6746-6754.

44. **Ghosh S, Setlow B, Wahome PG, Cowan AE, Plomp M, Malkin AJ, Setlow P.** 2008. Characterization of spores of *Bacillus subtilis* that lack most coat layers. Journal of bacteriology **190:**6741-6748.

45. **Schmidt S, Adolf F, Fuchsbauer H-L.** 2008. The transglutaminase activating metalloprotease inhibitor from *Streptomyces mobaraensis* is a glutamine and lysine donor substrate of the intrinsic transglutaminase. FEBS Letters **582:**3132-3138.

46. **Kashiwagi T, Yokoyama K, Ishikawa K, Ono K, Ejima D, Matsui H, Suzuki E.** 2002. Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. The Journal of biological chemistry **277:**44252-44260.

47. **Tagami U, Shimba N, Nakamura M, Yokoyama Ki, Suzuki Ei, Hirokawa T.** 2009. Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. Protein Engineering Design and Selection **22:**747-752.

48. **Yang MT, Chang CH, Wang JM, Wu TK, Wang YK, Chang CY, Li TT.** 2010. Crystal Structure and Inhibition Studies of Transglutaminase from *Streptomyces mobaraense*. Journal of Biological Chemistry **286:**7301-7307.

49. **Shimba N, Yokoyama K, Suzuki E.** 2002. NMR-based screening method for transglutaminases: rapid analysis of their substrate specificities and reaction rates. Journal of agricultural and food chemistry **50:**1330-1334.

50. **Hoffmann C, Schmidt G.** 2004. CNF and DNT. Rev Physiol Biochem Pharmacol **152:**49-63.

51. **Schmidt G, Selzer J, Lerm M, Aktories K.** 1998. The Rho-deamidating cytotoxic necrotizing factor 1 from *Escherichia coli* possesses transglutaminase activity. Cysteine 866 and histidine 881 are essential for enzyme activity. The Journal of biological chemistry **273:**13669-13674.

52. **Schmidt G, Goehring UM, Schirmer J, Lerm M, Aktories K.** 1999. Identification of the C-terminal part of *Bordetella* dermonecrotic toxin as a transglutaminase for rho GTPases. The Journal of biological chemistry **274:**31875-31881.

53. **Schmidt G, Goehring UM, Schirmer J, Uttenweiler-Joseph S, Wilm M, Lohmann M, Giese A, Schmalzing G, Aktories K.** 2001. Lysine and polyamines are substrates for transglutamination of Rho by the *Bordetella* dermonecrotic toxin. Infect Immun **69:**7663-7670.

54. **Buetow L, Flatau G, Chiu K, Boquet P, Ghosh P.** 2001. Structure of the Rho-activating domain of *Escherichia coli* cytotoxic necrotizing factor 1. Nature structural biology **8:**584-588.

55. **Makarova KS, Aravind L, Koonin EV.** 1999. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. Protein Sci **8:**1714-1719.

56. **Luo Y, Pfister P, Leisinger T, Wasserfallen A.** 2002. Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases produced in *Methanothermobacter* strains. FEMS microbiology letters **208:**47-51.

57. **King JD, Vinogradov E, Preston A, Li J, Maskell DJ.** 2009. Post-assembly modification of *Bordetella bronchiseptica* O polysaccharide by a novel periplasmic enzyme encoded by *wbmE*. The Journal of biological chemistry **284:**1474-1483.

58. **Milani A, Vecchietti D, Rusmini R, Bertoni G.** 2012. TgpA, a protein with a eukaryotic-like transglutaminase domain, plays a critical role in the viability of *Pseudomonas aeruginosa*. PLoS One **7:**e50323.

59. **Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM.** 2003. Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. The Journal of biological chemistry **278:**23834-23841.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt